Topics in Hedonic Valuation

by

Lala Xun Ma

Department of Economics
Duke University

Date: ____________________________

Approved:

____
Christopher Timmins, Supervisor

____
Peter Arcidiacono

____
Patrick Bayer

____
V. Joseph Hotz

____
Hugh Macartney

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Topics in Hedonic Valuation

by

Lala Xun Ma

Department of Economics
Duke University

Date: ________________________
Approved:

____
Christopher Timmins, Supervisor

____
Peter Arcidiacono

____
Patrick Bayer

____
V. Joseph Hotz

____
Hugh Macartney

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Environmental goods such as clean air and water are integral to human quality of life. However, because these amenities are not priced, their monetary values are not directly apparent. As a result, Hedonic methods have been employed as a tool to recover household Marginal Willingness To Pay (MWTP) for these goods to inform policy-making given constrained public resources. This thesis consists of three chapters tied to the Hedonic valuation of a particular environmental 'bad,' proximity to a brownfield site. Brownfield properties are lands that cannot be used due to the presence of a low-risk, hazardous substance.

The first chapter (joint work with Kevin Haninger and Christopher Timmins) uses property value hedonics to reveal household willingness to pay for brownfield cleanup . We exploit variation in space and time to deal with the potential bias in estimating MWTP due to unobservable variables that are correlated with both housing prices and site cleanup. Furthermore, there has been recent work showing that if equilibrium price functions change over time, the capitalization of changes in neighborhood amenities into property values over time (e.g. brownfield cleanup) may neither represent the preferences of those living in the neighborhood before changes occurred or after Kuminoff and Pope (2011). To address this, we provide a way to estimate cleanup without assuming that the hedonic price function is stable over time, an assumption that would likely be violated if site cleanup brought about significant changes to the community populations around the sites.

The second chapter considers two sources of distortions in the valuation of non-marketed goods - an expectations bias and a learning bias. If consumers suspect that cleanup of a brownfield is likely before it is cleaned (expectation) or gain new information about the severity of the brownfield contamination (information), then baseline period prices need to be adjusted to account for these potential distortions to the MWTP estimate. To address this, I collect a new data set on brownfield contamination information over time from Massachusetts, and recover hedonic values from a dynamic neighborhood choice framework that allows agents to learn about brownfield hazards in a Bayesian fashion. I find a MWTP estimate of $888.38 per unit of site contamination when accounting for learning and forward-looking behavior, which is more than double the simple hedonic estimate. Furthermore, parameters from my model can be used to calculate the average value of information provided by a site assessment.

The final chapter, joint work with Gabrielle Inder, examines whether different types of information about brownfield contamination capitalize into property values differently. More specifically, we estimate a property value hedonic model to test if housing prices are impacted differently if information about nearby contamination is released as a continuous measure as opposed to a binary measure (i.e. exceeding a threshold value or not). We do this by exploiting variation in contaminant thresholds used, holding constant the contaminant level, due to regulatory requirements for brownfield investigations in the State of Massachusetts. As the variation in threshold levels are tied to the level of human exposure of the areas in which these contaminated sites exist, threshold exceedance is potentially correlated with unobserved neighborhood characteristics that also impact housing values, creating the potential for biased parameter estimates. To deal with this, we take an instrumental variables approach using variation in threshold exceedance due to the location of underground water sources. After instrumenting for threshold exceedance with the presence of an

underground aquifer, our estimates indicate a 10.8% decrease in housing values from exceeding contaminant thresholds conditioning on contaminant levels, while continuous toxicity levels have a negative but insignificant effect. These findings suggest that polices aimed to improve public awareness about pollution should be cognizant of how information is conveyed, as it may allow for better design of information provision programs aimed to improve environmental quality.

To my mother,

who has sacrificed so much

so that I may be unconstrained

in my life's pursuits.

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my advisor, Christopher Timmins, for being encouraging and patient with me all of these years while providing me with invaluable research feedback at every step of the way. Chris has always looked out for opportunities that would benefit my career, short and long term. I sincerely cannot imagine having a different advisor and could not have asked for a better teacher, mentor, and role model.

My research has also benefited immensely from the various expertise of the other faculty on my committee: Peter Arcidiacono, Patrick Bayer, V. Joseph Hotz, and Hugh Macartney. In spite of having numerous other responsibilities, they have always been incredibly generous with their time, and for that I am grateful.

I wish to thank my co-authors, Kevin Haninger, from the U.S. Department of Health and Human Services for creating and giving me access to the brownfields data, and Gabrielle Inder, an undergraduate at Duke, for doing a great job collecting additional brownfields data. I am grateful to Yannis Ioannides from Tufts University for his guidance during my undergraduate years and for encouraging me to apply to doctorate programs in economics even after I had graduated from Tufts.

I would like to acknowledge Resources for the Future for generous funding under the Joseph L. Fisher Doctoral Dissertation Fellowship. In addition, the funding and administrative support from the Duke Economics Department was crucial to the completion of this dissertation.

Last but not least, I was lucky enough to have friends and classmates who were there for me to make the tough times bearable in the moment and bittersweet in retrospect. I am indebted to Lily Liu, Nixon Li, and Chutima (Gift) Tontarawongsa as I most likely would not have made it past that first year had they not spent countless nights during the qualifer summer answering my questions, discussing solutions, and making sure I was fed. I am always appreciative of my officemate, Ashley Vissing, who has seen the best and worst of me during the job market year, and of my college friend, Lien Duong, whose fortuitous choice of pursuing graduate studies one town over made her a continued source of emotional support. These individuals are, and will always remain, sources of knowledge and inspiration to me and, more importantly, close friends, whose love and encouragement have made my experience at Duke an enjoyable one.

1

Estimating the Impacts of Brownfield Remediation on Housing Property Values

with Kevin Haninger[1] *and Christopher Timmins*[2]

1.1 Introduction

Land revitalization is a beneficial, yet costly, process to undertake. Lands are often contaminated with various harmful substances that require expensive procedures to treat. In some instances, toxic waste sites are shown to pose a direct threat to human health. In other cases, sites may pose a low risk to nearby residents, but are left unused (or under-used) until even small amounts of contaminants are removed. Most would agree upon the importance of treating (or at least containing) health hazards at high-risk sites. As for low-risk sites, however, it is far less obvious that the benefits of remediation should exceed the costs. Even though these sites may not be especially toxic, their oftentimes poor aesthetic quality combined with their additional need for special treatment in order to be re-developed causes the surrounding area to be an undesirable place to live or work. Thus, the benefits of revitalizing these sites include

[1]Haninger: U.S. Department of Health and Human Services

[2]Timmins: Duke University Department of Economics and NBER

the economic development that would result from making them more productive and attractive. The EPA has designated these lower-risk sites as brownfields and has aimed to promote their revitalization through grant funding.

1.1.1 Identifying the Effects of Brownfield Remediation

This paper uses a slate of quasi-experimental approaches to estimate the benefits of brownfield cleanup by examining its effect on nearby property values. In this respect, the paper draws upon the extensive literature on property-value hedonics to recover homeowner willingness-to-pay for remediation.[3] The value of cleanup, as captured by the value capitalized into nearby housing prices, is a good way to measure a variety of beneficial effects, including effects on numerous local neighborhood amenities. Under certain conditions that we describe below, these capitalization effects can be given a welfare interpretation, making them particularly useful for cost-benefit analysis. In an ideal research environment, one would randomly select brownfield sites for cleanup and observe the impacts of that cleanup on nearby housing prices. The random selection of sites into the remediation process would guarantee that unobservable determinants of changes in local housing prices would not be correlated with changes induced by remediation, allowing the researcher to cleanly identify the latter. While more common in some areas of research, opportunities for these sorts of experiments are not often available in environmental economics.[4] Indeed, it is the case that the Brownfields Program awards cleanup grants based on a competitive process. The outcome of this process may lead to the award of cleanup funds to locations that differ systematically from locations that do not receive funds. To the extent that we can control for these differences with observables, they do not present a problem. Data describing sites and the neighborhoods around them are limited,

[3]See Taylor (2003) and Palmquist (2005) for summaries of this literature

[4]See Banerjee and Duflo (2009) for a description of the extensive role played by randomized experiments in development economics.

however, so there are necessarily going to be variables that we cannot control for directly.

We therefore adopt a variety of quasi-experimental approaches to identifying the effect of cleanup on brownfields. The idea of these approaches is to exploit some source of exogenous variation in data that approximates that which would result from a truly random experiment. We begin by demonstrating the bias that could result from ignoring unobservable confounders altogether with a cross-sectional specification. In particular, we compare locations with an untreated brownfield to areas with a remediated brownfield. The problem is that these groups may differ systematically with respect to unobservables that could be correlated with treatment status.

We then demonstrate how even a simple fixed effects specification, which uses changes in a neighborhood's exposure to an unremediated brownfield site, can help solve the problem. In particular, if unobservable differences between houses in the different neighborhoods are constant over time, we can difference that heterogeneity away by looking at changes in exposure status accompanying cleanup activities. Of course, only houses surrounding sites that are remediated experience a change in exposure status, so we must limit our analysis to houses in these neighborhoods.

The problem with the fixed effects specification is that not all unobserved factors will be time-invariant. If brownfield cleanup funds are typically awarded to 'up-and-coming' neighborhoods, the effect of cleanup will be confounded by those other improvements. The opposite would be true if awards were made in an attempt to turn around declining neighborhoods. Fixed effects are unable to deal with these time-varying unobservable factors that are correlated with cleanup activity. This is where we move to techniques traditionally considered 'quasi-experimental.'

First, we consider the 'difference-in-differences' (DID) specification. This approach defines a *treatment* group (e.g., the houses immediately surrounding a brownfield that is treated at some point in time t^*) and a *control* group (e.g., houses nearby to those in the treatment group, so that we can safely assume that other time-varying neighborhood factors will be the same, but far enough away so as to be able to assume that the impact of the brownfield site is negligible).[5] DID then compares the change in prices in the treatment group from houses sold in $t > t^*$ to those sold in $t < t^*$ to a similarly defined change in the control group. The change in prices in the control group, intuitively, controls for any changes in price induced by neighborhood-specific factors aside from brownfield remediation. The remaining effect can therefore be ascribed to the cleanup. Note in addition that, in the process of differencing within the treatment or the control groups, any time-invariant differences between these groups are controlled for as well.

The DID approach to estimation requires a number of non-trivial assumptions. The most important is the 'common trends' assumption - in particular, that the change over time in log price in the treatment and control groups would be the same (conditional upon observable covariates) were the treatment group to have remained untreated. This assumption is not testable. In addition to the common trends assumption, the DID specification requires that the equilibrium hedonic price function remain stable over time in order to give estimates a welfare interpretation. The same is also true of the fixed effects specification. We describe this issue in more detail in the following subsection, and use a DID matching estimator that avoids using time variation to deal with it.

1.1.2 Capitalization v. Marginal Willingness to Pay

The fixed effect and DID approaches to recovering the benefits of site remediation suffer from a similar problem. In particular, each requires an assumption that the hedonic price function, which describes the equilibrium relationship between house

[5] In practice, 'treatment' will consist of several stages, including assessment and cleanup activities, that we will model explicitly.

attributes (including exposure and treatment status) and price, is stable over time. Given the substantial neighborhood turnover that may occur in response to brownfield redevelopment, this assumption may be questionable. Put differently, with a new local population, the willingness-to-pay for not being exposed to an untreated brownfield site that is revealed by the hedonic price function may be very different after cleanup. Kuminoff and Pope (2011) show that the results of simple fixed effect estimation of the price response to cleanup may therefore fail to identify the MWTP of either those living in proximity to the brownfield before or after cleanup. Instead, it will recover a 'capitalization' effect (i.e., the simple response of price to a cleanup, without any additional welfare interpretations). The capitalization effect of a cleanup may be interesting in its own right (e.g., considering implications for property tax revenue collection), but it does not imply a welfare interpretation.

To overcome this problem, we suggest an alternative to using time variation under the traditional Diff-in-Diff estimator. In particular, we use a DID nearest-neighbor matching estimator (DD-NNM) that exploits the differences between both treatment and control groups within a neighborhood surrounding a particular site, and the differences between cleaned and uncleaned sites. This method compares similar houses in treatment and control groups around sites that were and were not cleaned, but does not require any comparisons over time. Matching of similar sites relies, in particular, on the state the brownfield is in, the number of previous assessments performed, the type of grant proposal (petroleum or hazardous substances), and Brownfield Program grant scores, which provide a good source of exogenous variation in cleanup status for otherwise similar sites.[6] By 'double differencing' in this manner (instead of over time), we are able to cleanly identify a different hedonic price function

[6] Applications receiving higher scores are more likely to be funded, but in any particular year a given score may or may not be funded owing to variability in the program's budget - simply put, the program works its way down the list of ranked proposals allocating funds until the budget runs out.

in each year. By not relying on time variation and an assumption of a stationary hedonic gradient, we are able to interpret our estimates as willingnesses to pay instead of simply capitalization effects.

Together, our fixed effect and quasi-experimental approaches to estimation all lead to a common conclusion - that cleanups conducted under the Brownfield Program yield a large, statistically significant, positive, but highly-localized effect on housing prices.

1.1.3 A Note on Localized Externalities

Brownfields, like many other disamenities (Superfund sites, TSDF's, TRI plants) may have very localized impacts on house prices. As such, it can prove difficult to recover these impacts without access to high-resolution data. Cleanup of a brownfield, for example, may not be perceptible in information about census tract median housing prices, while it may in fact have large impacts on nearby houses. One solution to this problem is to use high-resolution decennial census block-level data (Gamper-Rabindran et al. (2011)). That approach, however, introduces two potential problems. First, low-frequency decennial data may confound brownfield cleanup with other unobserved events that occurred at some other time during the same decade. Unlike Superfund remediation, brownfield cleanups can be relatively quick, leaving a great deal of remaining time over a ten-year period for other things to happen. Second, cleanups under the Brownfield Program have all taken place in the last decade, and long-form decennial census data have not been collected since 2000. These data are now collected as part of the American Community Survey, and are available at high geographic resolution only on a 'moving average' basis (e.g., for the period 2005-2009). Given that brownfield cleanup can be initiated and completed relatively quickly, we would not know whether most of the cleanups in our data set occurred before or after the homeowner valuations stated in the 2005-2009 ACS data.

In light of all of these concerns, we employ housing transactions data from Dataquick, Inc. that are both high-resolution (i.e., latitude and longitude) and high-frequency (i.e., day of transaction). This allows us to measure the impact of the cleanup with a great deal of precision, both in space and time.

1.1.4 Limitations of the Analysis

Before proceeding, we acknowledge a few limitations of our analysis. First, looking at the price of housing in close proximity to brownfield sites will not capture equilibrium effects that are realized elsewhere in the urban area - i.e., cleanup of brownfields may have impacts on local labor markets and on particular housing markets far from the brownfield in question. We will fail to capture these effects to the extent that they appear in other parts of the city. Given the size of a typical brownfield (relative to the size of an urban area), this may not be much of a practical issue. Still, we do note that new methods (i.e., estimable sorting models) may be able to deal with these sorts of concerns. (Kuminoff et al. (2013))

Second, our approach will also not capture health benefits from remediation that people are not aware of (and, hence, are not reflected in house purchase decisions and transactions prices). In contrast to other nuisances (Superfund sites, TSDF's, or other toxic waste exposure), we do not expect this to be as much of an issue for brownfield sites, making property value hedonics a good approach in this context.

1.1.5 Outline

This paper is divided into six sections. Section 2 describes the EPA Brownfields Program and cleanup process, paying particular attention to the cleanup grant application and scoring process. Section 3 describes our methodological approach, detailing the different specifications we use to recover estimates of MWTP in the presence of correlated unobservables. Section 4 describes the data, and Section 5 reports estimates from each specification. Section 6 concludes with a brief discussion and 'back-of-the-envelope' cost-benefit analysis.

1.2 The EPA Brownfields Program

A brownfield is a 'real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.'[7] Typically, brownfields are lands that were previously used for industrial or commercial purposes and include areas that are contaminated by low concentrations of hazardous substances. These sites are diverse in nature and can range from being old dry cleaning establishments and gas stations to processing plants for materials such as steel, bricks, and asbestos. Generally, brownfields pose lower risk to human health than other types of hazardous waste sites, as they exclude sites listed or proposed for listing on the National Priorities List and sites that are remediated under the Toxic Substances Control Act of 1976. The U.S. Government Accountability Office estimates that there are more than 450,000 brownfields nationwide. In 1995, the U.S. Environmental Protection Agency initiated the Brownfields Program to assist public and private sector organizations in revitalizing brownfields, mainly by providing grant funding. The aim was not only to improve the environment, but also to promote social and economic reinvestment in these unused lands. In 2002, the Small Business Liability Relief and Brownfields Revitalization Act (i.e., the 'Brownfields Law') was signed as an amendment to the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA), which established the Superfund program. The passage of the Brownfields Law formalized EPA policies regarding brownfields and expanded financial and technical assistance for brownfield remediation through the Brownfields Program.

[7] http://epa.gov/brownfields/. See the EPA's website for further details on the Brownfields Program and a link to public law 107-118 (H.R. 2869), 'Small Business Liability Relief and Brownfields Revitalization Act'.

1.2.1 Brownfield Grants

Brownfields grants serve as the foundation of the Brownfields Program and support land revitalization efforts by funding environmental site assessment, cleanup, and job training activities. There are four types of competitive grants that serve specific purposes in the land revitalization process. Assessment grants provide up to $200,000 for a grant recipient to inventory, characterize, assess, and conduct planning and community involvement related to brownfields sites. Job training grants provide funding to recruit mostly unemployed, low-income and minority residents from brownfield-affected areas and to train these individuals to secure full-time jobs in site assessment and cleanup. Cleanup grants provide up to $200,000 to perform cleanup activities at a brownfield site contaminated by petroleum and hazardous substances. Finally, revolving loan fund grants provide funding to capitalize a revolving loan fund, which is used to make loans and sub-grants for cleanup activities at brownfield sites. Since passage of the Brownfields Law through FY 2011, EPA has competitively awarded 1,479 assessment grants totaling $331.3 million, 143 revolving loan fund grants totaling $167.5 million, 801 cleanup grants totaling $150.7 million, and 121 job training grants totaling $25.2 million.

1.2.2 Cleanup Grant Applications, Proposal Scoring, and Awards

This paper focuses on the effect of cleanup grants on housing values. As stated above, cleanup grants provide up to $200,000 to perform cleanup activities at a brownfield site contaminated by petroleum or hazardous substances. Due to budgetary limitations, no eligible entity may apply for funding cleanup activities at more than three sites. Cleanup grants require a 20 percent cost share in the form of a contribution of money, labor, material, or services for eligible and allowable costs; however, applicants may request a waiver of the cost share requirement based on financial hardship. The performance period for cleanup grants is three years. Cleanup grant proposals are evaluated against both threshold and ranking criteria. Applicants must pass all threshold criteria in order to quality for funding. Threshold criteria include site ownership and eligibility for federal brownfield assistance, community notification and opportunity for public comment prior to proposal submission, and a letter from the appropriate state or tribal environmental authority acknowledging that the applicant plans to apply for federal brownfield assistance.

Conditional upon passing all threshold criteria, the proposal will receive a numerical score from the evaluation panel. Scores are based on several evaluation fields, including community need, project description and feasibility, community involvement and partnerships, and reduction of threats to human health and the environment. Once scored, cleanup grant proposals are ranked from highest to lowest score and then awarded funding in rank order until the program budget has been exhausted.[8]

If a proposal is not awarded in one year, the applicant can reapply in a subsequent year. Within the universe of brownfield cleanup proposals, we identified 172 properties that reapplied for funding at least once in the six-year period after the program began, 87 of which was eventually awarded funding. This implies that the brownfield site could be associated with different proposal scores and different award statuses. We take the applicant's most recent score and application outcome, assuming that it represents the applicant's best and most knowledgeable proposal effort. More details on how scores are compared across grant years are provided in Section 4.

1.3 Model and Identification

Since brownfield cleanup activity is not directly traded in markets, a revealed preference approach is used to infer its value from its impact on nearby housing prices.

[8] More information on the cleanup grant application process can be found at http://www.epa.gov/brownfields/applicat.htm.

This paper uses the hedonic method to model a property's price.[9] For a thorough discussion of the hedonic method, see the reviews by Taylor (2003) and Palmquist (2005). The hedonic price function is defined as a mapping from the attributes of a house, including the presence of a nearby brownfield, to a price in equilibrium. The implicit price of brownfield exposure may be measured with, for example, the hedonic price gradient with respect to distance.

The hedonic method is based on the idea that homeowners' disutility from living in close proximity to a brownfield site can be measured by observing compensating price differentials in housing markets. In general, the homeowner's marginal willingness to pay (MWTP) for some desirable attribute (e.g., distance from a brownfield site) can be read off of the hedonic gradient (i.e., the derivative of the hedonic price function), owing to utility-maximizing homeowners' sorting behavior. Rosen's seminal paper (Rosen (1974)) and the literature it sparked describe procedures for recovering the MWTP functions for heterogeneous individuals. Bishop and Timmins (2011) describe many of the difficulties encountered in this exercise - because of these difficulties, the typical approach in the applied hedonics literature has been to ignore this heterogeneity and either recover a function that describes price as a linear function of distance, or one that treats exposure discretely, defining it according to whether a house falls inside a particular distance band drawn around a brownfield. That is the approach we adopt here.

One of the more difficult problems that arises when implementing the hedonic method is the presence of house and neighborhood attributes that are unobserved by the researcher but correlated with the attribute of interest. These unobservables have the potential to bias the results of a simple cross-sectional specification. Em-

[9] Assuming that the housing supply is fixed in the short-run, any improvement to a brownfield is assumed to be completely capitalized into price and not in the quantity of housing supplied. Given that the Brownfields Program is relatively recent, we would expect to still be in the 'short-run'. As more time passes, researchers will be able to study whether cleanups have had a discernible impact on new development.

pirical approaches that are used to deal with this problem include (i) fixed effects, (ii) differences-in-differences, and (iii) matching estimators. We briefly review the econometric theory behind each of these modeling strategies below.

1.3.1 Cross-Sectional Estimates

The simplest specification ignores any panel variation in the data. Considering all houses in counties that contain brownfields[10], the most straightforward comparison is to compare houses in the vicinity of a brownfield site to ones in the county that are not exposed to a site at all. Exposure is defined geographically; houses located inside a circular buffer of 5km surrounding a site are considered to be 'exposed'. However, houses and neighborhoods near brownfields are likely to be different in unobservable ways from those that are not, and these unobservables may lead to biased estimates. Table 1.1 describes the observable attributes of houses surrounding brownfields in our sample compared with those not exposed to brownfields in the rest of the county, regardless of cleanup status (but before any cleanup has occurred at sites that are cleaned). A simple inspection of this table suggests several reasons to be concerned about the results of a simple cross-sectional analysis. In particular, there are statistically and economically significant differences between houses that lie in close proximity (5km) to a brownfield and those that do not - e.g., houses within 5km of a brownfield site are less expensive and tend to be older and smaller than those in the rest of the county. These large differences in observables suggest that there may also be differences in unobservable attributes of each of these groups of sites.

An alternative approach limits the analysis to only houses surrounding brownfields within 5km of sites (both those that have and those that have not been cleaned). By limiting the sample in this way, we narrow the variation in unobservable hetero-

[10]We describe the sample of brownfield sites we use for estimation in Section 4.

geneity that might be correlated with brownfield exposure. We estimate the following regression specification:

$$P_i = \beta_0 + \beta_1 CLEANUP_i + X_i'\delta + YEAR_i'\gamma + \epsilon_i \tag{1.1}$$

where

P_i = log of transaction price of house i

$CLEANUP_i = 1$ if brownfield that house i is exposed to has been treated under the Brownfields Program[11]

X_i = vector of attributes of house i

$YEAR_i$ = vector of dummy variables indicating year in which house i is sold

The effect of cleanup is then measured by β_1. The problem here is that $CLEANUP_i$ is still likely to be correlated with ϵ_i. Potential bias arises if brownfields that received treatment were systematically different in unobservable ways from those that did not receive treatment. We might, for example, expect that houses located in close proximity to awarded brownfields may be of lower quality than those located near non-awarded sites. Table 1.2 compares houses surrounding cleaned brownfields from our sample to those surrounding brownfields that have not been cleaned. While Table 1.2 shows that the size of those differences is dramatically lower than are the differences between areas with and without brownfield sites, evidence of significant differences between houses lying inside a 5km buffer of sites that are eventually cleaned compared with those that are not eventually cleaned still exist. For subsequent methods, we limit the analysis to only using houses within 5km of brownfield sites.

[11]In practice, this will be houses exposed to awarded brownfields *after* the brownfield has undergone cleanup.

1.3.2 Fixed Effects

The simplest approach to dealing with unobserved house and neighborhood attributes that may be correlated with brownfield remediation is to exploit the variation in panel data to control for time-invariant neighborhood attributes. Suppose P_{itk} measures the natural log of the price of house i located in the neighborhood around brownfield k which transacts in year t. X_{itk} is a vector of attributes of that house,[12] and $CLEANUP_{itk}$ is a dummy variable that takes the value 1 if the brownfield k has completed the cleanup process by period t (= 0 otherwise).[13] As in equation 1.1, consider only houses that are in close proximity to brownfields.[14] μ_k is a time-invariant attribute associated with the neighborhood around brownfield site k that may or may not be observable by the researcher, and ν_{itk} is a time-varying unobservable attribute associated with the house. Importantly, μ_k may be correlated with $CLEANUP_{itk}$ (i.e., sites that receive cleanup treatment may be in neighborhoods that are systematically different from those that do not receive cleanup).

$$P_{itk} = \beta_0 + \beta_1 CLEANUP_{itk} + X'_{itk}\delta + \mu_k + \nu_{itk} \tag{1.2}$$

[12] Note that, with Dataquick data, house attributes do not vary over time because only the structural attributes from the most recent property assessment is recorded. We subscript X by k and t simply to indicate the neighborhood in which the house is found and the year in which it transacts.

[13] Housing transactions observed before the start of the cleanup period are given a value of $CLEANUP_{itk} = 0$. In practice, we distinguish between houses sold before and during cleanup from those sold after. We discuss this in detail in Section 4.

[14] We present estimates from using multiple buffers to demonstrate robustness.

Using $(i,k) \in k$ to denote all houses in all years that lie in the neighborhood surrounding site k, we can take the within-neighborhood means of each variable:

$$\begin{aligned}
\overline{P_k} &= \frac{1}{N_k} \sum_{(i,k)\in k} P_{itk} \\
\overline{CLEANUP_k} &= \frac{1}{N_k} \sum_{(i,k)\in k} CLEANUP_{itk} \\
\overline{X_k} &= \frac{1}{N_k} \sum_{(i,k)\in k} X_{itk} \\
\overline{\mu_k} &= \frac{1}{N_k} \sum_{(i,k)\in k} \mu_{itk} \\
\overline{\nu_k} &= \frac{1}{N_k} \sum_{(i,k)\in k} \nu_{itk}
\end{aligned} \tag{1.3}$$

and generate mean-differenced data:

$$\begin{aligned}
\tilde{P}_k &= P_{itk} - \overline{P_k} \\
CLE\tilde{A}NUP_k &= CLEANUP_{itk} - \overline{CLEANUP_k} \\
\tilde{X}_k &= X_{itk} - \overline{X_k} \\
\tilde{\mu}_k &= \mu_{itk} - \overline{\mu_k} \\
\tilde{\nu}_k &= \nu_{itk} - \overline{\nu_k}
\end{aligned} \tag{1.4}$$

Noting that $\mu_{itk} - \overline{\mu_k} = 0$, we can then re-write equation 1.2:

$$\tilde{P}_{itk} = \beta_1 CLE\tilde{A}NUP_{itk} + \tilde{X}'_{itk}\delta + \tilde{\nu}_{itk} \tag{1.5}$$

Estimating this specification therefore controls for any permanent unobservable differences between places that received cleanup treatment and those that did not.

1.3.3 Difference-in-Differences (DID)

Let P_{itk} be the log of the price of house i in the neighborhood surrounding brownfield site k at time t. At some point in time, site k is cleaned. Considering only houses in the vicinity of brownfields that are cleaned (5km), and let the treatment group of houses be defined by those that are close enough (i.e. closer than 5km) to be affected by that cleanup. A specific definition of treatment is discussed in section 3.4, but the intuition is that these houses are particularly close to the brownfield, while there may be other houses in the same local neighborhood that experience the same local public goods but are far enough from the brownfield to not be 'treated' by it. We define this distance below.

The dummy variable $TREAT_{ik}$ is equal to 1 if house i belongs to the treatment group (i.e., is located within some buffer b, less than 5km, surrounding the brownfield), and it is equal to 0 if it belongs to the control group (i.e., inside 5km but outside the treatment group). Let $POST_{tk}$ indicate post-treatment, which equals 1 if a house lying within 5km of site k (in either the treatment or control group) sells after site k is cleaned. The model for the observed log price is then written as

$$P_{itk} = \beta_0 + \beta_1 TREAT_{ik} + \beta_2 POST_{itk} + \pi TREAT_{ik} \times POST_{itk} + u_{itk} \tag{1.6}$$

where π represents the expected change in log price for the treated group less the expected change in price for the control group. According to the above model, it is equal to

$$\begin{aligned}\pi = \Big(& E\left[P_{itk} \mid TREAT_{ik} = 1, POST_{itk} = 1\right] \\ & - E\left[P_{itk} \mid TREAT_{ik} = 1, POST_{itk} = 0\right]\Big) \\ -\Big(& E\left[P_{itk} \mid TREAT_{ik} = 0, POST_{itk} = 1\right] \\ & - E\left[P_{itk} \mid TREAT_{ik} = 0, POST_{itk} = 0\right]\Big)\end{aligned} \tag{1.7}$$

Using the 'potential outcomes' notation (Rubin 1974), where P^0_{itk} represents the log of i's potential price if the house does not receive treatment and P^1_{itk} represents the log of i's potential price if it does.

$$\begin{aligned}\pi = &\Big(E\left[P^1_{i1k} \mid TREAT_{ik} = 1\right] - E\left[P^0_{i0k} \mid TREAT_{ik} = 1\right]\Big) \\ &- \Big(E\left[P^0_{i1k} \mid TREAT_{ik} = 0\right] - E\left[P^0_{0} \mid TREAT_{ik} = 0\right]\Big)\end{aligned} \tag{1.8}$$

The main identifying assumption underlying the DID model is that of common trends, which specifies that

$$\begin{aligned}&E\left[P^0_{i1k} \mid TREAT_{ik} = 1\right] - E\left[P^0_{i0k} \mid TREAT_{ik} = 1\right] \\ &\qquad = E\left[P^0_{i1k} \mid TREAT_{ik} = 0\right] - E\left[P^0_{i0k} \mid TREAT_{ik} = 0\right]\end{aligned} \tag{1.9}$$

In the case of brownfields, this assumption implies that, in the absence of cleanup, the potential log prices of properties in the treated group would have followed the same trend as log prices in the control group. Under this assumption, π identifies the Average Treatment Effect on the Treated (ATT). In particular, we can use equation 1.9 to replace the third and fourth terms in equation 1.8:

$$\begin{aligned}\pi = &\Big(E\left[P^1_{i1k} \mid TREAT_{ik} = 1\right] - E\left[P^0_{i0k} \mid TREAT_{ik} = 1\right]\Big) \\ &- \underbrace{\Big(E\left[P^0_{i1k} \mid TREAT_{ik} = 0\right] - E\left[P^0_{i0k} \mid TREAT_{ik} = 0\right]\Big)}_{E\left[P^0_{i1k} | TREAT_{ik}=1\right] - E\left[P^0_{i0k} | TREAT_{ik}=1\right]}\end{aligned} \tag{1.10}$$

Canceling repeated terms yields

$$\pi = E\left[P^1_{i1k} \mid TREAT_{ik} = 1\right] - E\left[P^0_{i1k} \mid TREAT_{ik} = 1\right] \tag{1.11}$$

Failing to control for observable covariates may invalidate the common trends assumption. One can easily control for them by extending the regression model used to recover π:

$$P_{itk} = \beta_0 + \beta_1 TREAT_{ik} + \beta_2 POST_{itk} + \pi TREAT_{ik} \times POST_{itk} + X'_{ik}\delta + u_{itk} \tag{1.12}$$

In practice this regression model can be expanded to include multiple groups and multiple treatment periods. For application to brownfield cleanup, we separate the pre-cleanup time frame into two periods and make all comparisons to prices before cleanup activities begin. This will be elaborated in Section 4.1.

1.3.4 Defining Treatment and Control Groups

The DID specification allows one to control for two types of unobservables. First, it controls for unobservables that vary by group (treatment and control) but not over time. Second, it controls for unobservables that affect outcomes over time but are common to both groups. This motivates the definition of treatment and control groups to identify cleanup impact. One way is to define the treatment group as the properties near a brownfield that is cleaned and the control group as those near a brownfield that remains uncleaned. However, if the two brownfields are located in different places, it is likely that the prices of surrounding houses will be subject to unobservables that are not only group-specific, but which also change over time. An example would be if brownfields that are cleaned are in up-and-coming neighborhoods compared to those that are not. Over time, the prices of houses near cleaned brownfields would reflect this improvement, compromising the DID identification strategy.

Instead of defining treatment and control groups as above, this paper follows the strategy employed by Linden and Rockoff (2008), using adjacent neighborhoods around a brownfield to define treatment and control groups to alleviate the problem of group- and time-specific unobservables.[15] That is, houses located within a certain distance of a brownfield are considered to be in the treatment group, while houses located outside of that distance (where the site has no effect regardless of cleanup) are designated as controls. To find that distance, we estimate two functions describing

[15]Linden and Rockoff (2008) estimate the impact of sex offender arrival in Mecklenberg County, North Carolina.

the relationship between price and the distance to the nearest brownfield for all property transactions occurring before and after cleanup. Ideally, the distance at which the difference in the price functions becomes insignificant is the point at which we would define the cutoff between the treatment and control groups.

Specifically, one would expect that prices of properties closer to brownfield sites are impacted more by cleanup than those located far away. Furthermore, at some distance far enough away from the site, cleanup should not influence property prices at all. It has been found that the effects of hazardous waste sites such as those on the National Priorities List decrease very quickly with distance from the site (Adler et al. (1982), Kohlhase (1991), Kiel (1995)). This suggests that the treatment and control groups can be defined by the distance at which brownfields begin to have no impact. If this were the case, then price shocks that would affect the trend of one group would arguably affect that of the other group as well. Ultimately, the common trend assumption is untestable. However, this paper provides graphical evidence in the data section and specification tests in the results section that allow us to better assess the validity of this assumption.

1.3.5 MWTP v. Capitalization

The intention when running the hedonic specifications described above is to recover an estimate of the MWTP for the amenity in question (here, cleanup of a proximate brownfield site). Kuminoff and Pope (2011) note that price function estimates identified using changes in prices and amenities over time formally recover a capitalization rate (i.e., the rate at which housing prices increase with the change in the amenity). This may not be the same as the MWTP (i.e., the actual slope of the hedonic price function) either before or after the amenity change. Moreover, it is hard to say a priori which direction the difference between the capitalization effect and the MWTP might go. As long as the hedonic price function is constant over time, there should not be a difference between capitalization and MWTP. One would therefore expect the difference between MWTP and capitalization to be smaller the shorter is the time-period between observations.

To make this point clear, consider the simple example of two hedonic gradients that apply to two different time periods (indexed by $t = 1, 2$):

$$P_{1k} = \rho_1 + \theta_1 g_{1k} + \mu_k + \epsilon_{1k} \tag{1.13}$$

$$P_{2k} = \rho_2 + \theta_2 g_{2k} + \mu_k + \epsilon_{2k}$$

In this example, g_{tk} indicates the policy being valued.[16] The MWTP in each period is given by θ_1 and θ_2, respectively. If we were to take the difference between these two equations in order to eliminate the fixed effect, μ_k, we would obtain:

$$\Delta P_k = (\rho_2 - \rho_1) + (\theta_2 g_{1k} - \theta_1 g_{1k}) + \Delta\epsilon_k \tag{1.14}$$

However, estimating this equation requires the (stronger than usual) assumption that both g_{1k} and g_{2k} are uncorrelated with $\Delta\epsilon_k$. As such, we typically assume $\theta_1 = \theta_2 = \phi$ and instead estimate

$$\Delta P_k = \psi + \phi\Delta g_k + \Delta\epsilon_k \tag{1.15}$$

where $\psi = \rho_2 - \rho_1$. To see how this may yield biased estimates of both θ_1 and θ_2 , note the following:

$$\phi = \frac{B - A}{g_2 - g_1} \qquad A = \rho_1 + \theta_1 g_1 \qquad B = \rho_2 + \theta_2 g_2 \tag{1.16}$$

[16] Applied to the question of brownfield remediation, P_{tk} might refer to the log of the median price in the neighborhood surrounding brownfield site k and $g_{t,k}$ would refer to whether that site has been cleaned by period t. In our estimates, we allow for house-level variation in the transaction price data. For the purposes of illustrating Kuminoff and Pope's point, however, it is simpler to describe the model estimated using only site-level variation.

Therefore,

$$\begin{aligned}
\phi &= \frac{\theta_2 g_2 - \theta_1 g_1}{g_2 - g_1} + \frac{\rho_2 - \rho_1}{g_2 - g_1} \qquad (1.17)\\
&= \frac{\theta_2 g_2 - \theta_1 g_1}{g_2 - g_1} + \left(\frac{\theta_2 g_1}{g_2 - g_1} - \frac{\theta_2 g_1}{g_2 - g_1}\right) + \frac{\rho_2 - \rho_1}{g_2 - g_1}\\
&= \frac{g_1\left(\theta_2 - \theta_1\right)}{g_2 - g_1} + \theta_2 + \frac{\rho_2 - \rho_1}{g_2 - g_1}
\end{aligned}$$

It is therefore easy to see that, if $\theta_2 = \theta_1$, ϕ will recover the common MWTP estimate. However, if this is not the case, there is no reason why ϕ even has to lie inside the range defined by θ_1 and θ_2 .

In the previous two sub-sections, we discussed estimators where the distinction between capitalization and MWTP is a potential issue. While we can take some comfort in the fact that we are typically relying on variation in prices over just a few years (and, hence, the hedonic price function may not have much time to evolve), we propose a strategy that deals explicitly with this problem in the following sub-section. In particular, we estimate a separate hedonic price function in each year by exploiting variation in data across treated houses in cleaned and uncleaned sites.

1.3.6 Difference-in-Differences Nearest Neighbor Matching (DD-NNM)

We begin this sub-section by returning to the specification used to estimate the difference-in-differences model in sub-section 3.3, but allowing all of the parameters of the hedonic price function to vary with time. Furthermore, we index each observation by i (house), t (year) and k (site near to which house i is located). Some of the sites have been cleaned by time t ($CLEANUP_{tk} = 1$) while others have not ($CLEANUP_{tk} = 0$). Note that we include the set of sites that applied for, but were denied funding (i.e., $CLEANUP_{tk} = 0 \; \forall \; t$). Finally, we include a flexible function of house attributes (h). We consider only transactions that occur in a particular year t; we therefore do not need to differentiate between a pre- and post-treatment periods. Instead, we only need to differentiate between sites that have and have not been cleaned:

$$\begin{aligned} P_{itk} = \beta_{0t} + \beta_{1t} TREAT_{ik} + \beta_2 CLEANUP_{tk} + \\ \pi_t TREAT_{ik} \times CLEANUP_{tk} + f(h_{itk}; \theta_t) + u_{itk} \end{aligned} \tag{1.18}$$

We begin by considering only houses in a particular year t that are inside the treatment buffers of either a cleaned or an uncleaned site. As such, $TREAT_{ik} = 1$ for all houses in this sample,

$$P_{itk} = (\beta_{0t} + \beta_{1t}) + (\beta_2 + \pi_t) CLEANUP_{tk} + f(h_{itk}; \theta_t) + u_{itk} \tag{1.19}$$

Using a nearest-neighbor matching algorithm, we pair each house inside the treatment buffer in each neighborhood with $CLEANUP_{tk} = 1$ with a set of J houses that are as similar as possible in h_{itk} and located inside the treatment buffer of a neighborhood with $CLEANUP_{tk} = 0$. We also match on the sites' average proposal scores, proposal type, whether there was a phase II assessment performed, and restrict matches to be between sites in the same state.

Specifically, for a particular house i located in the treatment buffer of a cleaned site (price designated by P_{itk}), we find the J = 5 'nearest neighbors' to i, t, k (prices denoted by $P_j^{(itk)}$.

$$(\beta_{2t} + \pi_t) = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(P_{itk} - \frac{1}{J} \sum_j P_j^{(itk)} \right) \tag{1.20}$$

Next, we repeat this process using only those houses transacted in year t that are located outside the treatment buffer (but inside 5km) in neighborhoods surrounding sites that were not cleaned (i.e., $TREAT_{ik} = 0$ for all of these houses). Denoting the prices of houses located outside the treatment buffer with a $\tilde{P}_{itk}$, we get:

$$\beta_{2t} = \frac{1}{\tilde{N}_t} \sum_{i=1}^{\tilde{N}_t} \left(\tilde{P}_{itk} - \frac{1}{J} \sum_j \tilde{P}_j^{(itk)} \right) \tag{1.21}$$

As such, we are able to recover an estimate of the treatment effect on the treated for each year t by calculating:

$$\pi_t = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(P_{itk} - \frac{1}{J} \sum_j P_j^{(itk)} \right) - \frac{1}{\tilde{N}_t} \sum_{i=1}^{\tilde{N}_t} \left(\tilde{P}_{itk} - \frac{1}{J} \sum_j \tilde{P}_j^{(itk)} \right) \tag{1.22}$$

The success of this strategy, of course, depends upon being able to find high-quality matches for houses in neighborhoods around cleaned sites from the set of houses around sites that have not been cleaned. This is what assures that the unspecified function $f(h_{itk}; \theta_t)$ will be differenced away. By limiting comparisons to houses within 5km of sites, and by matching based on brownfield characteristics and restricting matches to be amongst sites in the same state, moreover, we eliminate other forms of heterogeneity at the neighborhood level.

1.4 Data

Our analysis is based on two main sources of data. In the following three subsections, we describe the data, define our pre- and post-treatment periods, and provide summary statistics along with graphical evidence supporting our identification assumptions.

1.4.1 Data Description

Data on brownfield properties are provided by the Environmental Protection Agency. The data set includes all brownfields that applied for cleanup grants in the years 2002 through 2008. The data provide characteristics of the brownfields, including the exact location (latitude and longitude),[17] property size (for awarded sites only), and types of grant application (i.e. targeted to treat petroleum sites, sites with hazardous substances, or both). A subset of this applicant pool is awarded a cleanup grant. Generally, one brownfield is tied to one cleanup grant, although there are a few cases where a brownfield is tied to multiple grants. For the properties that were awarded funding, the data include related award and cleanup progress information. Since funding for brownfields varies each year and is awarded beginning with the highest scoring applicant and working downward until funding runs out, there is not one score cutoff that determines whether a property is cleaned. Moreover, because of changing scoring rules, the raw scores are difficult to compare across competition years. To make scores comparable across years, we standardize the scores to be between 0 and 100 by dividing the raw score by the maximum score in its respective competition year.

Dates of different milestones in the process to remediate the brownfield exist starting from site assessment and ending with cleanup. However, these dates are not always available for all of the awarded and non-awarded sites, so we consider all houses sold before any cleanup activities commence to belong to a period, 'Pre-cleanup.' Next, we define an interim treatment period that starts from the earliest recorded cleanup start date, and ends on the cleanup completion date.[18] We distinguish this interim period as houses sold during this time are not exposed to the full effect of cleanup. Lastly, we define the post-cleanup period during which properties have been fully treated with brownfield cleanup as starting with the cleanup completion date and lasting for the duration of our sample.

[17]Available information describes the centroid of the brownfield property, but not property boundaries. This is a common feature in data describing the geographic siting of locally undesirable land uses (i.e., LULU's). Like most of this literature, we use distance from the centroid as a measure of exposure. Obtaining more detailed information that would allow us to measure the distance to a site's boundary would be desirable.

[18]Dates on which information are released to the public about cleanup, such as the public announcement of grant awards, are also reasonable to consider.

The time period dummy variables that will be used in all of the specifications are $Interim_{tk}$ and $Post_{tk}$, which respectively equal to 1 if a house is sold during and after cleanup of the nearby brownfield. For the DID specification, interactions between each of the above time period dummies with the treatment dummy are included. In that specification, the coefficient on $Treat_{tk} \times Post_{tk}$ is the treatment effect on the treated, and should be interpreted with respect to the houses in the pre-cleanup period, which is the omitted time period. There are several brownfields where cleanup activities have not begun or are not yet complete. We retain these brownfields for the analysis. As long as the types of brownfields that are awarded earlier in our sample (and importantly the types of neighborhoods in which they are situated) do not differ from those that are awarded later on, this should not affect our estimates.

The second data source comes from Dataquick Information Systems, used under a license agreement with the Duke Department of Economics, which provides housing transactions data. These data contain the history of transactions and characteristics for houses in a large number of U.S. counties. The data include information on the sale of newly constructed houses, re-sales, refinance or equity dealings, timeshare sales, and subdivision sales. The data saves transaction-related information such as price, date and associated loans. For each house in the data set, the attributes are recorded from the most recent tax assessment. The attribute fields are detailed and include characteristics such as the number of bedrooms, bathrooms, square footage, lot size, number of units, and number of stories. The housing assessment data also include the latitude and longitude of each property.

In addition to house-level attributes, we control for county level effective real estate tax (RET) rate (Siniavskaia (2011)), as defined by the percentage of the property value that is paid in taxes every year. The county-level RET rates are calculated using homeowner-reported home values and annual real estate taxes from the Census Bureau's 2005-2009 American Community Survey.[19]

The set of brownfields under consideration are those tied to cleanup grant applications between 2002 and 2008. There are a total of 1383 brownfield applications in the EPA data, 446 of which are awarded cleanup grants and 937 are not. Applicants could reapply for a grant in another year following a rejection. Taking into consideration re-applications, we identified 1178 unique brownfield properties. After removing brownfields with missing or inaccurate longitude and latitude coordinates, we are left with 797 sites (437 awarded and 360 non-awarded). Property locations were individually verified with Google maps and checked to ensure that the background of the reported location corroborated with the information from the grant proposal. Dataquick does not have housing data for all counties in which brownfields are located; therefore, only a subset of the properties that are tied to cleanup grants are included. Out of a total of 797 unique brownfields from the EPA data with geocoordinates, 327 had associated housing transactions data within 5 kilometers of only 1 brownfield site. Of those 327 sites, 197 are awarded with cleanup and 130 are not.[20] Currently, the window of observations used for housing transactions starts in 1998 (four years before the start of the brownfields program)[21] and ends in 2012, which is the last available year for housing sales.

Focusing on the housing data, our analysis limits transactions to house sales or re-sales of owner occupied properties. Houses with missing prices, bathrooms, bedrooms, or square footage are dropped. Furthermore, since only housing characteristics from the most recent tax assessment are recorded, any house indicated to have undergone major improvements is dropped, as its attributes may be incorrect for previous transactions. To reduce possible errors in record-keeping and sales anoma-

[19]For details, see Siniavskaia (2011).

[20]This figure is after removing certain locations where house attributes are missing.

[21]The extent of geographic coverage by Dataquick becomes much greater in 1998. Going back further in time would require dropping more brownfield sites for lack of housing data.

lies, the analysis excludes houses that sold more than once per year or five times in the eleven year window of house sales.[22] Prices are normalized to January 2000 dollars using the monthly, regional All Urban Consumers Housing CPI taken from the Bureau of Labor Statistics. The analysis excludes the 1st and 99th percentile of the observed price distribution.

Knowing the exact locations of all properties allows us to calculate the distance between each house and the nearest brownfield. This is our measure of brownfield 'exposure'. Using Graphical Information Systems (GIS), each property is first matched to the nearest brownfield within a 5 kilometer radius. The distances to those brownfields are then recovered. Houses not within 5 kilometers of any brownfield are dropped. Houses located near multiple brownfields, in which case the effect of cleanup may be hard to measure, are dropped. The treatment and control groups are then defined using houses within this 5 kilometer radius. Even though the houses outside of 5 kilometers will not be used in the estimation, it is of interest to compare differences between close (within 5 kilometers) and to those located in the rest of the county (in addition to comparing treatment and control houses within 5 kilometers) in order to motivate the employed definition of treatment. We define both treatment and control groups to be contained in a small area around brownfield sites (5km) to minimize the threat of any location-specific unobservable differences that may affect price dynamics.

An important note is that the available EPA data describe the set of brownfield sites associated with applications for cleanup grants. This precludes analysis of brownfields that did not apply for funding. Therefore, it is possible that there are brownfields (along with other locally undesirable land uses) in neighborhoods that are not accounted for. Even though the analysis cannot control for these sites, it

[22]The former often represent non-arms-length transactions that can sometimes lead to multiple transactions on the same day. The latter (i.e., more than 5 transactions in 11 years) signals that the house may be used as an investment property by a house 'flipper' (Bayer et al. (2011b)).

is unlikely that the status of these brownfields will have changed over the course of our analysis, making them time-invariant unobservables that will be differenced out of our analysis using several of the methods described in the previous section. Moreover, if they do change status over time, our DID estimator will control for this to the extent that they equally affect treatment and control groups.

1.4.2 Graphical Evidence

The next step is to determine the distance at which the control and treatment groups are defined. We begin by estimating a pair of price functions over distance from the nearest brownfield site - one for pre-cleanup transactions and one for post-cleanup transactions. The distance at which the pre-cleanup and post-cleanup price functions converge is where brownfield cleanup no longer impacts house prices; this is ideally where we would define the cutoff between treatment and control groups.

Rather than impose a functional form for the price function, we use a local linear polynomial estimator (Fan and Gijbels (1996)), which is described in detail in the appendix.[23] We make one modification to this procedure to account for the fact that the mix of houses sold before and after cleanup changes with respect to distance. In particular, Figure 1.1 describes the average square footage of houses sold at each distance from a brownfield site before and after cleanup. It is clear from this figure that houses sold before cleanup of brownfield sites within approximately 2km tend to be larger than those sold in that same buffer after cleanup. We therefore control parametrically for house attributes before recovering the non-parametric relationship between house prices and distance in Figure 1.2. Figure 1.2 also controls parametrically for year effects to allow for general inflationary trends, differences in brownfield characteristics including the proposal scores, proposal type, and the number of times

[23]The bandwidth, determined by inspection, is three times Silverman's Rule of Thumb. For the distance gradient, this is about 308 meters. For the time gradient, it is approximately 381 days. A Gaussian kernel is used for weighting.

the sites are assessed.[24]

Figure 1.2 provides evidence in support of the assumption that houses that are 'far' enough from the brownfield represent a valid control group. While we find that houses at all distances have higher prices on average after cleanup, we find that this difference narrows outside of 2030 meters. Taking the treatment group to be defined by a 2030m buffer, the simple DID estimator will compare the average change in prices before assessment and after cleanup inside the buffer with the similarly defined change outside the buffer. We demonstrate the sensitivity of some of our results to the assumed buffer size in the following section.

Given the definition of the treatment and control groups, a natural way to check whether the common trend assumption is reasonable is to compare the price trends of the treatment and control groups pre- and post-treatment. If the common trend assumption is valid, then price trends should exhibit a few characteristics. First, if the relationship between price and cleanup is causal, one would expect a significant price increase for treatment houses around the time of cleanup, as opposed to a gradual upward trend in price. This would support the claim that cleanup in fact leads to an increase in prices of houses near brownfields. Second, the price trends of the two groups in the pre-cleanup period should be relatively similar (i.e., common trends before cleanup). Third, in the post-cleanup period, the prices of the control houses should not change significantly, but rather should follow a path similar to that in the pre-treatment period. The latter two characteristics would suggest that price trends for houses near brownfields would have been the same as those far from brownfields had they not been treated with cleanup.

Figure 1.3 plots the prices of treatment (i.e., inside 2030m) and control houses

[24] All brownfields must undergo Phase I and II site assessments. Under certain circumstances, however, additional testing may be advised by a Licensed Site Professional, and a supplemental site assessment is conducted. Recognizing those sites that demand additional testing may control for differences in the severity of contamination at sites.

against time relative to the cleanup date.[25] The trends pre- and post-treatment are similar for the two groups. While both groups exhibit a jump at the point of treatment, suggesting that some of the treatment may spill-out into the control group, the discontinuity for the control group going from pre- to post-cleanup (-0.33%) is smaller than that in the treatment group (6.55%). The differences-in-differences approach measures the jump in the treatment group relative to that in the control group.

1.4.3 Summary Statistics

Table 1.3 provides summary statistics for the brownfields in the sample. The table provides statistics for subsets of brownfields by housing data availability in order to examine the representativeness of the sample after data cuts and merges. Columns (1) - (3) and (4) - (6), respectively, summarize characteristics of the subsets of brownfields with and without Dataquick housing data. Tests for the equality of group means for the various attributes across these subsets are provided in columns (7) and (8). Table 1.3 suggests that proposal scores are marginally higher for non-funded brownfields in locations with Dataquick data, compared to non-funded brownfields in locations without Dataquick data. The difference is not statistically significant for the set of funded properties. Hazardous substances contamination is more common in the funded brownfields for which we do not have housing data; since Dataquick does not provide data for many rural communities, significant differences may reflect the more common occurrence of certain types of brownfields in more urbanized areas.

Table 1.4 provides summary statistics for house attributes by treatment status. Columns (1) - (2) and (3) - (4), respectively, summarize the housing characteristics for the treatment group (within 2030m of a brownfield) and the control group (between

[25] As was the case when generating Figure 1.2, we parametrically control for housing attributes, year effects, and brownfield characteristics before non-parametrically estimating price as a function of time relative to the cleanup period.

2030m and 5km of a brownfield). Columns (5) and (6) test for equality of group means. Although we reject the equality of means for many attributes, we do take comfort in the fact that the differences are far smaller than in Table 1.1, which compares houses within 5km to houses in the rest of the county. We take Table 1.4 as evidence that there are important differences between treatment and control groups that should be accounted for parametrically in the DID specification.

Table 1.5 provides a yearly breakdown of cleanup starts and completions for the brownfields that were awarded cleanup grant funding.[26] Since the Brownfields Law was only recently enacted in 2002, many cleanup completions occur towards the end of the window of observations, which limits the number of post-cleanup transactions we have to work with. Table 1.6 reports the mean cleanup duration by toxin-found and media of contamination. The average cleanup duration for all brownfields for which we can calculate durations is approximately 444 days with a standard deviation of 451 days. These figures imply that brownfield cleanups are relatively quick (e.g., in comparison to the cleanup of a Superfund site); this requires that we use high-frequency housing data (i.e., daily transactions information) for estimation.

Even with the relatively short average duration of brownfield cleanup, right-censoring (i.e., cleanups that are not completed by the end of our sample) is still an issue - particularly for cleanups begun in later years. Table 1.7 describes the fraction of cleanups initiated in each year that were not completed by 2011.[27] Not surprisingly, cleanups begun later in the sample are less likely to be completed. There is, however, a significant fraction of cleanups with petroleum contamination begun early in the sample that have not been completed by 2012.

We find that being near a brownfield site that has been cleaned yields prices

[26]There are 2 sites that began cleanup before the 2002 - one for areas with Dataquick coverage, and without. These are likely from pilot programs that receive funding before the formal program began.

[27]There were no cleanups initiated in 2012 from the pool of awarded sites between 2002 and 2008.

that are consistently lower than houses that are not near sites (by -3.23% to -9.11%, depending upon the buffer size), and lower than values of houses that are near brownfield sites in any other state of cleanup activity. The coefficients for BF indicates that prices are marginally lower for houses located near an untreated site.

1.5 Empirical Results

1.5.1 Cross-Sectional Estimates

Table 1.8 reports the results of our cross-sectional specification described in equation 1.1, where we restrict the comparison to be between houses that are in the vicinity of brownfields - some of which have been cleaned, others of which have not. We find that the value of cleanup is negative at -11.3%. The counterintuitive sign of this effect may be a result of omitted variables bias if cleanup grants are targeted towards struggling neighborhoods. Table 1.8 suggests that unobservable neighborhood attributes may be correlated with their cleanup status, necessitating a different empirical approach.

1.5.2 Fixed Effect Estimates

Next, we use the fixed effects specification described in equation 1.5, which controls for time-invariant unobservables associated with neighborhoods. These unobservables can be the source of bias that leads to the counterintuitive results found in the cross-sectional specifications. The fixed effects specification uses all houses in a buffer; we consider buffers of 1000, 2000m, 3000m, and 5000m to demonstrate robustness. We also include controls for year fixed effects, house attributes, and the real estate tax rate. The results of the fixed effects specification, described in Table 1.9, differ strikingly from the cross-sectional results, with increases in house prices from cleanup that range between 6.24% and 11.1%, depending on the size of the buffer.[28]

[28] Only the Fixed Effect estimate using a buffer at 2000 meters is significant using cluster-robust standard errors.

1.5.3 Difference-in-Differences Estimates (DID)

While it is able to deal with time-invariant unobservable neighborhood attributes, the fixed effects specification described in Table 10 does nothing to control for time-varying unobservables that may be correlated with brownfield cleanup. Estimates would still be biased if, for example, cleanup were systematically directed towards locations that were improving in unobservable ways. The DID approach overcomes this problem with the 'common trends' assumption - namely, that the change over time in unobservables in the control group is the same as it would have been in the treatment group in the absence of treatment. By assigning the control group to be houses in the same neighborhood as those in the treatment group, but far enough away from the site to not be impacted by cleanup, we try to satisfy this assumption and obtain estimates that account for any time-varying unobservables that are common to both the treatment and control groups. Moreover, by differencing over time, the DID approach also controls for time-invariant unobservables just as the fixed effects specification did.

As described in Section 3, the average treatment effect on the treated is measured by the coefficient on the interaction of the indicators for a house being in the treatment group ($Treat$) and its transaction occurring after the cleanup has been completed ($Post$). These estimates can be found in the fifth row of Table 1.10. With only year fixed effects and brownfield-level controls, we find a treatment effect of 5.81% using the preferred buffer size of 2030m. In a specification that includes year fixed effects, house-level and brownfield-level controls, and controls for the real estate tax, this effect increases to 7.15%. Further introducing brownfield fixed effects decreases this effect to 4.85%, which is significant at the 10% after clustering standard errors at the brownfield level. In the specification with brownfield fixed effects, the estimate for the cleanup interim interaction additionally reveals that cleanup might slightly depress housing values compared to the houses in the same area that are located outside of the exposure buffer, suggesting that cleanup effort, though on average fairly quick, can be disruptive.

1.5.4 Difference-in-Differences Nearest Neighbor Matching Estimates (DD-NNM)

Both the fixed effects and DID approaches rely on the strong assumption that the hedonic price function remains stable over time. If cleanup activities initiate neighborhood turnover, the identities of those living in close proximity to the site may change, and with them, marginal willingness to pay may change as well. In fact, Kuminoff and Pope (2011) demonstrate that estimates of the hedonic price function may provide no information about MWTP. As such, one needs a method that both controls for unobservables that may be correlated with cleanup activities while not relying on time variation. The difference in differences nearest neighbor matching estimator described in Section 3 is designed to do this.

Estimates of the average treatment effect on the treated (π) are recovered without using time variation by taking the difference between two sets of parameter estimates - one derived by comparing houses inside the treatment buffer of cleaned sites to houses inside the treatment buffers of uncleaned sites ($\beta_2 + \pi$), and the other derived by comparing houses in the control groups of cleaned sites to houses in the control groups of uncleaned sites (β_2). Table 1.11 describes these estimates for our preferred buffer size of 2030 using J = 5 matches. Estimates and standard errors are based on Abadie and Imbens (2006).

We do not consider results for 2004, since there is not enough post cleanup observations for estimation. For 2005 and 2007, results based on inside-buffer comparisons are not statistically significant, but results are significant for both comparisons in other years. In particular, we find cleanup effects of 16.1% to 37.7%. Since our aim is to estimate the benefits from cleaning up brownfield sites, we also consider limiting the post-cleanup period to end at most 3 years after the cleanup completion date in Table 1.12, as brownfield cleanup can trigger other types of neighborhood redevelopment activities that are not directly tied to the cleanup itself. Using at most 3 years after cleanup, the DD-NNM estimator finds effects of 10.7% to 24.8%.[29] These results suggest that we can indeed interpret our results as implying a positive and significant willingness to pay for brownfield remediation (i.e., a welfare interpretation). The fixed effects estimator, which estimates an 11.1% increase in housing values, is comparable to the smallest of the DD-NNM estimator. Compared with the results of the fixed effects and DID specifications, these larger estimates suggest that changes in the price function over time may have indeed had the effect of reducing the estimated MWTP.

1.6 Cost-Benefit Analysis

Finally, we can address the simple question, 'is brownfield remediation worth it?' In answering this question, we take a conservative approach. First, we take our most conservative estimate of the cleanup effect - the difference-in-differences estimate based on a 2030m treatment buffer (4.85%), rather than the larger estimates generated by the fixed effect and DD-NNM specifications. Next, we take a conservative estimate of the value of housing that sold inside the treatment buffer prior to cleanup. Ideally, we would like to measure the total value of all housing units inside each buffer prior to the start of cleanup, but we do not observe every house sell during that pre-cleanup period. Rather than try to impute values for houses that we do not see transact during that period, we take the conservative approach of aggregating the value of only the houses that do sell in the 5 years prior to the start of cleanup inside the treatment buffer. We are able to construct this aggregate value

[29] Estimates are fairly robust to using matches of different sizes (See Tables 1.13 and 1.14 for estimates using J=10 matches).

for 51 of the brownfields - \$4,019,624,960. Multiplying by 4.85% yields an estimate of the aggregate increase in housing value owing to cleanup of \$194,951,811. This represents a benefit of aggregate value per site of \$3,822,585. If the \$200,000 EPA cleanup grant represented just 1/10th of the total cleanup cost, brownfield remediation would still pass a cost-benefit analysis. This result would be even stronger if we considered all the properties located inside the treatment buffer, a larger treatment buffer, or one of our larger treatment effect estimates.

1.7 Discussion

The EPA's Brownfields Program provides grants to assess and cleanup properties the 'expansion, re-development, or re-use of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant.' In this paper, we quantify the benefits associated with these remediation activities using property value hedonic techniques. As is typically the case in property value hedonic applications, omitted neighborhood attributes have the potential to bias these estimates. Indeed, our evidence suggests that neighborhoods that successfully cleanup brownfields under the program may be worse in other unobserved dimensions.

We offer a slate of quasi-experimental approaches to overcome this problem, including simple neighborhood fixed effects, a difference-in-differences approach that relies on a treatment and control group defined by geographic proximity, and a difference-in-differences nearest neighbor matching estimator that exploits the advantages of our treatment and control group definitions while not requiring that the hedonic price function remain stable over time. These alternative specifications yield a consistent conclusion - averaging over the experiences at a large number of brownfield properties, cleanup leads to housing price increases between 4.9% and 24.8%. Taking the most conservative estimate of the value of an average site cleanup, we find that it indeed passes cost-benefit analysis by an order of magnitude based on the expenditures from the Brownfields program. Moreover, our estimate using a difference-in-differences matching estimator without time variation is consistent with a willingness to pay (i.e., welfare) interpretation, not simply a capitalization effect. Aside from the empirical application in Kuminoff and Pope (2011), this is the only other empirical evidence that demonstrates the Kuminoff and Pope (2011) result. It suggests that caution must be exercised for assuming time-constant hedonic price functions in policy evaluation if the policy under consideration induces large enough changes such that the population considered before treatment is inherently different from the one after.

1.8 Tables

Table 1.1: House Attributes by within 5km versus Rest of County

	Within 5km		Rest of County		Equality of Means	
Attributes	Mean	St. Dev.	Mean	St. Dev.	t-stat	Reject?
Price	230,426.57	1,377,013.13	263,210.91	1,266,378.88	37.04	Y
Real Estate Rate (County)	10.94	4.78	9.90	4.98	-307.13	Y
Age	55.78	34.85	31.62	27.98	-1,141.92	Y
Square Footage	1,495.73	6,482.25	1,814.20	19,809.96	26.33	Y
Bathrooms	1.70	6.85	1.96	2.09	98.56	Y
Bedrooms	2.30	1.98	2.18	1.82	-88.12	Y
Sold in Year Built	0.03	0.16	0.07	0.25	251.08	Y
Condominium	0.19	0.39	0.16	0.37	-96.81	Y
Multifamily	0.09	0.28	0.02	0.13	-549.31	Y
Single Family	0.71	0.45	0.81	0.39	340.42	Y
Mobile	0.00	0.05	0.01	0.09	113.32	Y
Misc.	0.01	0.08	0.00	0.06	-45.89	Y
Obs.	2,769,158		9,288,332			

Note: Compares all houses within 5000m of a brownfield (funded or unfunded) before cleanup to houses located outside 5000m in the rest of the county.

Table 1.2: House Attributes by Whether Brownfield is Funded or Unfunded

	Funded Brownfields		Unfunded Brownfields		Equality of Means	
Attributes	Mean	St. Dev.	Mean	St. Dev.	t-stat	Reject?
Price	197,596.68	140,223.30	228,864.38	153,341.36	82.56	Y
Real Estate Rate (County)	10.51	4.01	9.74	4.86	-66.55	Y
Age	46.47	32.54	45.76	30.95	-8.76	Y
Square Footage	1,560.11	690.50	1,551.65	652.17	-4.97	Y
Bathrooms	1.87	0.81	1.99	0.83	56.55	Y
Bedrooms	3.01	1.05	3.03	1.01	6.42	Y
Sold in Year Built	0.04	0.20	0.03	0.18	-18.85	Y
Condominium	0.15	0.35	0.16	0.37	15.28	Y
Multifamily	0.05	0.22	0.04	0.19	-22.74	Y
Single Family	0.80	0.40	0.79	0.40	-6.48	Y
Mobile	0.00	0.04	0.00	0.06	13.32	Y
Misc.	0.00	0.03	0.00	0.06	17.12	Y
Obs.	250,809		395,756			

Note: Sample includes all houses located within 5000m of a brownfield (funded or unfunded). For funded brownfields, attributes are taken from houses selling before cleanup.

Table 1.3: Brownfield Attributes by Availability of Housing Data

	With Dataquick Data			Without Dataquick Data				
Variable	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.	t-stat	Reject?
Funded and Unfunded								
Petroleum	0.28	0.45	301	0.17	0.37	714	4.00	Y
Hazardous Substances	0.74	0.44	301	0.86	0.35	714	-4.48	Y
Proposal Score (std.)	76.29	12.74	301	74.73	11.46	449	1.75	Y
Funded Only								
Petroleum	0.32	0.47	139	0.16	0.37	138	3.24	Y
Hazardous Substances	0.69	0.46	139	0.88	0.33	138	-3.85	Y
Proposal Score (std.)	84.45	4.53	139	84.67	4.81	138	-0.40	N
Property Size (acres)	8.61	22.38	139	11.05	27.08	138	-0.82	N
Ready for Reuse	0.74	0.44	139	0.64	0.48	138	1.74	Y

Table 1.4: Housing Attributes by Treatment Status (Determined by Buffer)

	Treat ($\leq$ 2030m)		Control (>2030m)			
Attributes	Mean	St. Dev.	Mean	St. Dev.	t-stat	Reject?
Price	192,525.09	130,231.39	198,691.77	142,265.28	8.42	Y
Real Estate Rate (County)	9.96	3.62	10.63	4.08	32.43	Y
Age	52.95	36.00	45.07	31.57	-46.55	Y
Square Footage	1,560.49	690.42	1,560.03	690.52	-0.13	N
Bathrooms	1.83	0.78	1.88	0.81	11.15	Y
Bedrooms	3.10	1.19	3.00	1.01	-18.62	Y
Sold in Year Built	0.04	0.19	0.05	0.21	8.28	Y
Condominium	0.13	0.34	0.15	0.36	8.91	Y
Multifamily	0.09	0.29	0.04	0.20	-44.02	Y
Single Family	0.77	0.42	0.81	0.40	15.68	Y
Mobile	0.00	0.03	0.00	0.04	4.17	Y
Misc.	0.00	0.04	0.00	0.03	-1.54	N
Obs.	44,539		206,270			

Note: All Houses located within 5km of a brownfield (awarded only). Attributes are taken from houses selling before cleanup.

Table 1.5: Timeline of Brownfield Start and Completion Frequencies

	With Dataquick Data		Without Dataquick Data	
	Starts	Completions	Starts	Completions
2000	1			
2001			1	
2002	1		1	
2003	4		2	
2004	23	6	17	5
2005	37	12	23	13
2006	35	35	36	18
2007	23	24	28	26
2008	30	17	34	27
2009	30	23	17	33
2010	8	22	8	14
2011	1		1	2

Table 1.6: Brownfield Cleanup Duration (in Days) by Contaminant

Contaminant Funding Type	Mean	St. Dev.	Obs.
Petroleum only	444.08	468.75	60
Hazardous Substances only	442.72	449.64	210
Contaminant Found			
Controlled Substances	741.90	645.86	10
Asbestos	493.62	476.18	86
PCBs	489.58	468.92	45
VOCs	501.88	464.46	108
Lead	445.65	415.13	156
Other Metals	438.97	436.78	117
PAHs	448.07	436.83	117
Other	495.85	520.38	75
Unknown	383.00	513.36	2
Media of Contamination	Mean	sd	N
Soil	464.06	450.36	234
Air	329.33	289.09	12
Surface Water	356.00	282.92	21
Groundwater	520.44	499.20	126
Drinking Water	634.00		1
Sediments	422.45	441.28	11
Unknown	456.00	359.05	7

Table 1.7: Fraction of Proposals Initiated in Each Year (column) that Did Not Complete Cleanup by 2009

Contaminant Funding Type	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Petroleum only		0.0000	0.1765	0.0952	0.2857	0.3333	0.4444	0.6000		
Hazardous Substances only	0.0000	0.0000	0.0455	0.0541	0.1818	0.1905	0.2222	0.4865	0.4000	1.0000
		0.0000	0.0000	0.0000	0.0000		0.0000		1.0000	
Contaminant Found	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Controlled Substances			0.0000	0.0000	0.0000	0.0000		0.0000		
Asbestos	0.0000	0.0000	0.0769	0.0714	0.0000	0.1364	0.3684	0.5714	0.5000	
PCBs		0.0000	0.1667	0.0833	0.1000	0.2500	0.4444	0.5000	0.4286	1.0000
VOCs		0.0000	0.1429	0.1429	0.2667	0.3000	0.2286	0.5556	0.5000	1.0000
Lead		0.0000	0.0500	0.0345	0.1250	0.1563	0.2250	0.5517	0.4167	1.0000
Other Metals		0.0000	0.1538	0.0400	0.2258	0.2609	0.2963	0.5000	0.3636	1.0000
PAHs	0.0000	0.0000	0.0769	0.0800	0.1818	0.2000	0.2647	0.6800	0.4286	1.0000
Other	0.0000	0.0000	0.0000	0.0000	0.1333	0.2857	0.0000	0.3000	0.3333	
Unknown			0.0000				0.0000			
Media of Contamination	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
Soil	0.0000	0.0000	0.1143	0.0600	0.2131	0.2222	0.2857	0.5610	0.4000	1.0000
Air			0.2500	0.0000	0.0000			0.0000		
Surface Water			0.2500	0.1667	0.1429	0.0000	0.0000	0.0000	0.5000	
Groundwater		0.0000	0.0909	0.1071	0.3030	0.3182	0.2895	0.6154	0.5455	1.0000
Drinking Water				0.0000		1.0000	1.0000			
Sediments			0.3333	0.2500	0.5000	0.0000	0.7500	0.5000	0.5000	
Unknown			0.0000	0.0000		0.3333	0.0000	0.5000	1.0000	

Note: No cleanups were initiated in 2012 that submitted applications for the 2002-2008 fiscal years.

Table 1.8: Cross-Sectional Specification

Comparison of Houses Near Cleaned versus
Not Cleaned Brownfields (within 5km)

VARIABLES	All within 5km
Cleanedup	-0.113***
	(0.00275)
Constant	11.52***
	(0.140)
Obs.	469,928
R-squared	0.481
Controls	
Year Fixed Effects	x
Brownfield Characteristics	x
House Controls	x
BF Fixed Effects	

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sample used includes only houses inside buffer 5 km around a funded brownfield that has been cleaned or an unfunded brownfield. Cleanedup = 1 if the house is near a funded site has been cleaned.

Table 1.9: Fixed Effects

VARIABLES	b= 1000	b= 2000	b= 3000	b= 5000
Interim	0.164***	0.223***	0.223***	0.209***
	(0.0572)	(0.0527)	(0.0525)	(0.0446)
Post	0.0816	0.111*	0.0975	0.0624
	(0.0716)	(0.0644)	(0.0631)	(0.0521)
Constant			10.57***	11.19***
			(0.262)	(0.188)
Obs.	18,686	64,652	136,480	370,910
Number of Brownfields	59	70	86	197
Controls				
Year Fixed Effects	x	x	x	x
Brownfield Characteristics				
House Controls	x	x	x	x
BF Fixed Effects	x	x	x	x

Note: Cluster-Robust Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample used includes only houses (i) around awarded brownfields, and (ii) inside buffer b in meters. Post = 1 if transaction occurs after nearby brownfield is cleaned. Interim = 1 if transaction occurs during cleanup.

Table 1.10: Differences-in-Differences (b=2030m)

VARIABLES			
Treat	-0.0477***	-0.0492***	-0.0458***
	(0.004)	(0.003)	(0.017)
Interim	0.237***	0.114***	0.104***
	(0.004)	(0.003)	(0.039)
Post	0.149***	0.00116	0.0749**
	(0.004)	(0.003)	(0.037)
Interim × Treat	-0.0799***	-0.0379***	-0.0191
	(0.009)	(0.007)	(0.023)
Post × Treat	0.0581***	0.0715***	0.0485*
	(0.007)	(0.006)	(0.029)
Constant	13.97***	13.40***	11.23***
	(0.022)	(0.199)	(0.169)
Obs.	370,910	370,910	370,910
R-squared	0.087	0.471	0.380
Number of Brownfields			197
Controls			
Year Fixed Effects	X	X	X
Brownfield Characteristics	X	X	
House Controls		X	X
BF Fixed Effects			X

Note: Cluster-Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Treatment buffer = 2030m. Sample used includes only houses (i) around awarded brownfields, and (ii) inside 5km buffer. Treat = 1 if house is within b buffer in meters. Post = 1 if transaction occurs after nearby brownfield is cleaned. Interim = 1 if transaction occurs during cleanup.

Table 1.11: Difference-in-Differences Nearest Neighbor Matching Estimator (b=2030, 5 Matches)

	Inside Treatment Buffer				Outside Treatment Buffer				Average Treatment Effect on the Treated
B=2030	# Matches	Est.	S.E.	Obs.	# Matches	Est.	S.E.	Obs.	
Y2005	5	-0.0271	(0.0353)	9,309	5	-0.297***	(0.0259)	31,089	26.99%
Y2006	5	-0.191***	(0.0262)	8,991	5	-0.389***	(0.0213)	31,047	19.80%
Y2007	5	-0.0439	(0.0292)	7,268	5	-0.205***	(0.0196)	26,626	16.11%
Y2008	5	-0.143***	(0.0320)	5,970	5	-0.343***	(0.0242)	21,989	20.00%
Y2009	5	-0.0778**	(0.0321)	6,509	5	-0.369***	(0.0230)	25,436	29.12%
Y2010	5	-0.223***	(0.0345)	6,537	5	-0.561***	(0.0229)	25,500	33.80%
Y2011	5	-0.270***	(0.0398)	6,068	5	-0.545***	(0.0225)	23,416	27.50%
Y2012	5	-0.120***	(0.0425)	4,995	5	-0.497***	(0.0248)	19,199	37.70%

Note: Standard errors in parentheses and calculated according Abadie and Imbens (2006). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.12: Double Difference Matching Estimator (b=2030, 5 Matches, 3 Years after Cleanup)

	Inside Treatment Buffer				Outside Treatment Buffer				Average Treatment Effect on the Treated
B=2030	# Matches	Est.	S.E.	Obs.	# Matches	Est.	S.E.	Obs.	
Y2005	5	-0.0271	(0.0353)	9,309	5	-0.297***	(0.0259)	31,089	26.99%
Y2006	5	-0.191***	(0.0262)	8,991	5	-0.389***	(0.0213)	31,047	19.80%
Y2007	5	-0.0412	(0.0289)	7,263	5	-0.210***	(0.0196)	26,564	16.88%
Y2008	5	-0.268***	(0.0355)	5,452	5	-0.497***	(0.0254)	20,047	22.90%
Y2009	5	-0.453***	(0.0449)	5,339	5	-0.639***	(0.0251)	20,621	18.60%
Y2010	5	-0.656***	(0.0642)	5,225	5	-0.763***	(0.0282)	19,544	10.70%
Y2011	5	-0.521***	(0.0687)	4,736	5	-0.769***	(0.0324)	17,150	24.80%
Y2012	5	-0.486***	(0.0768)	3,690	5	-0.710***	(0.0362)	13,419	22.40%

Note: Standard errors in parentheses and calculated according Abadie and Imbens (2006). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.13: Double-Difference Matching Estimator (b = 2030, 10 Matches)

	Inside Treatment Buffer				Outside Treatment Buffer				Average
									Treatment Effect
B=2030	# Matches	Est.	S.E.	Obs.	# Matches	Est.	S.E.	Obs.	on the Treated
Y2005	10	0.0354	(0.0293)	9,309	10	-0.139***	(0.0224)	31,089	11.19%
Y2006	10	-0.165***	(0.0220)	8,991	10	-0.338***	(0.0174)	31,047	14.70%
Y2007	10	-0.0597**	(0.0248)	7,268	10	-0.219***	(0.0169)	26,626	17.51%
Y2008	10	-0.118***	(0.0279)	5,970	10	-0.341***	(0.0249)	21,989	19.80%
Y2009	10	-0.0881***	(0.0280)	6,509	10	-0.372***	(0.0201)	25,436	29.42%
Y2010	10	-0.242***	(0.0292)	6,537	10	-0.531***	(0.0200)	25,500	30.80%
Y2011	10	-0.264***	(0.0324)	6,068	10	-0.555***	(0.0202)	23,416	28.50%
Y2012	10	-0.129***	(0.0357)	4,995	10	-0.474***	(0.0227)	19,199	35.40%

Note: Standard errors in parentheses and calculated according Abadie and Imbens (2006). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.14: Double-Difference Matching Estimator (b = 2030, 10 Matches, 3 Years after Cleanup)

	Inside Treatment Buffer				Outside Treatment Buffer				Average Treatment Effect on the Treated
B=2030	# Matches	Est.	S.E.	Obs.	# Matches	Est.	S.E.	Obs.	
Y2005	10	0.0354	(0.0293)	9,309	10	-0.139***	(0.0224)	31,089	17.44%
Y2006	10	-0.165***	(0.0220)	8,991	10	-0.338***	(0.0174)	31,047	17.30%
Y2007	10	-0.0578**	(0.0246)	7,263	10	-0.222***	(0.0168)	26,564	16.42%
Y2008	10	-0.279***	(0.0308)	5,452	10	-0.499***	(0.0203)	20,047	22.00%
Y2009	10	-0.376***	(0.0390)	5,339	10	-0.624***	(0.0222)	20,621	24.80%
Y2010	10	-0.619***	(0.0514)	5,225	10	-0.730***	(0.0248)	19,544	11.10%
Y2011	10	-0.520***	(0.0496)	4,736	10	-0.780***	(0.0278)	17,150	26.00%
Y2012	10	-0.486***	(0.0640)	3,690	10	-0.707***	(0.0322)	13,419	22.10%

Note: Standard errors in parentheses and calculated according Abadie and Imbens (2006). *** p<0.01, ** p<0.05, * p<0.1.

1.9 Figures

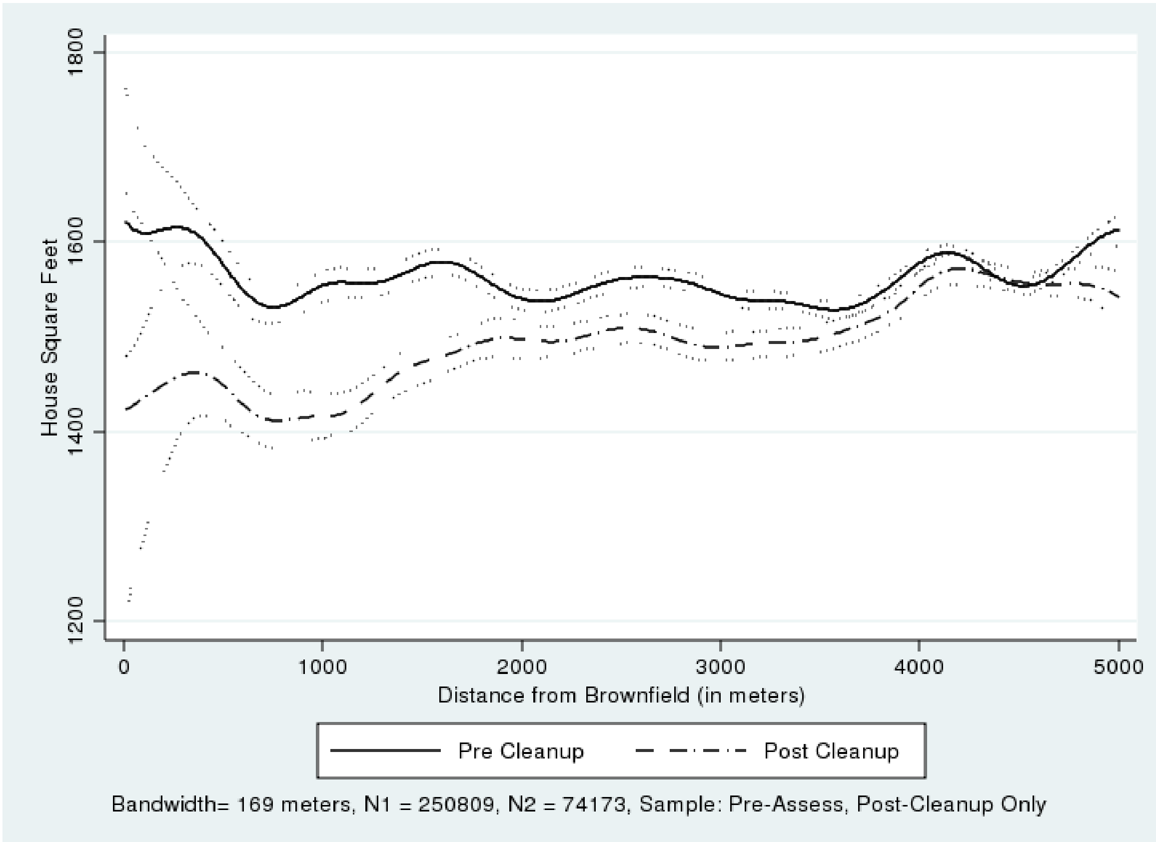


FIGURE 1.1: Average Square Footage of Houses Transacted by Distance from Brownfield Before v. After Remediation With 99% Confidence Intervals

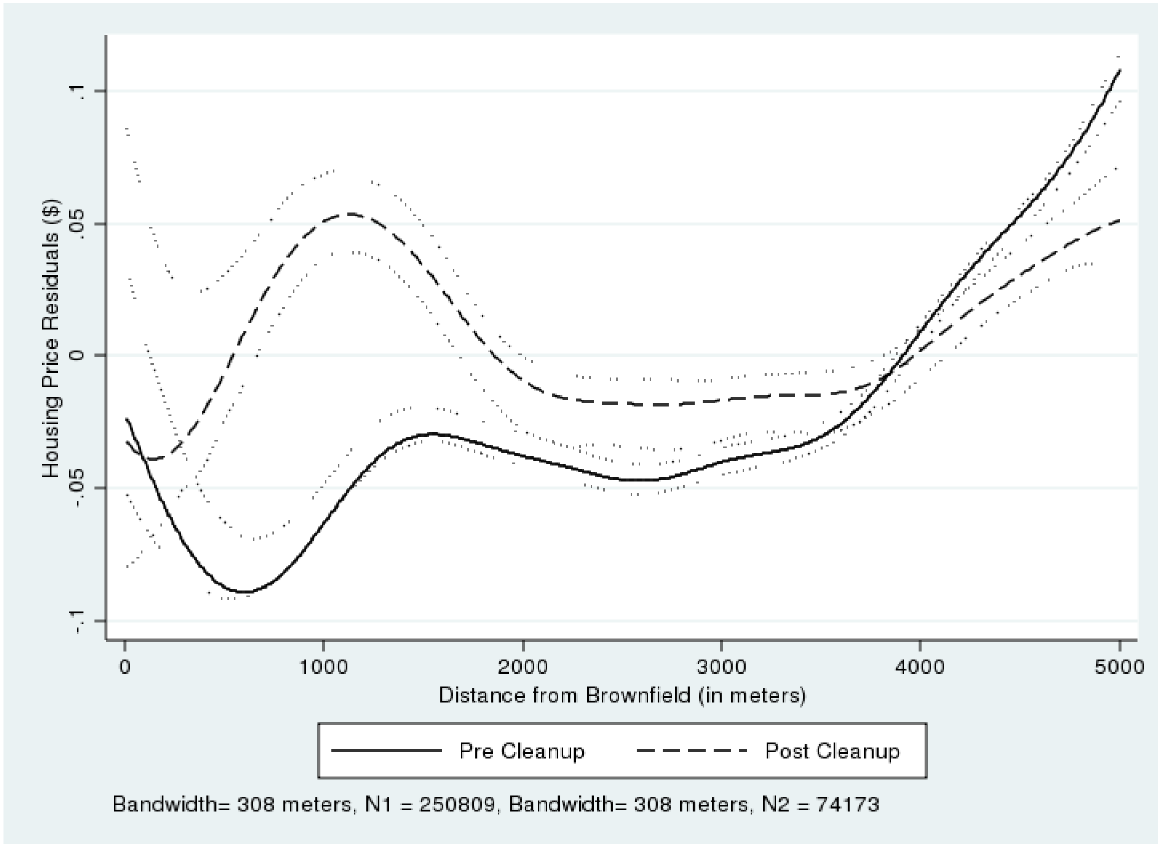


FIGURE 1.2: Non-Parametric Price Function Estimates Before and After Remediation With 99% Confidence Intervals

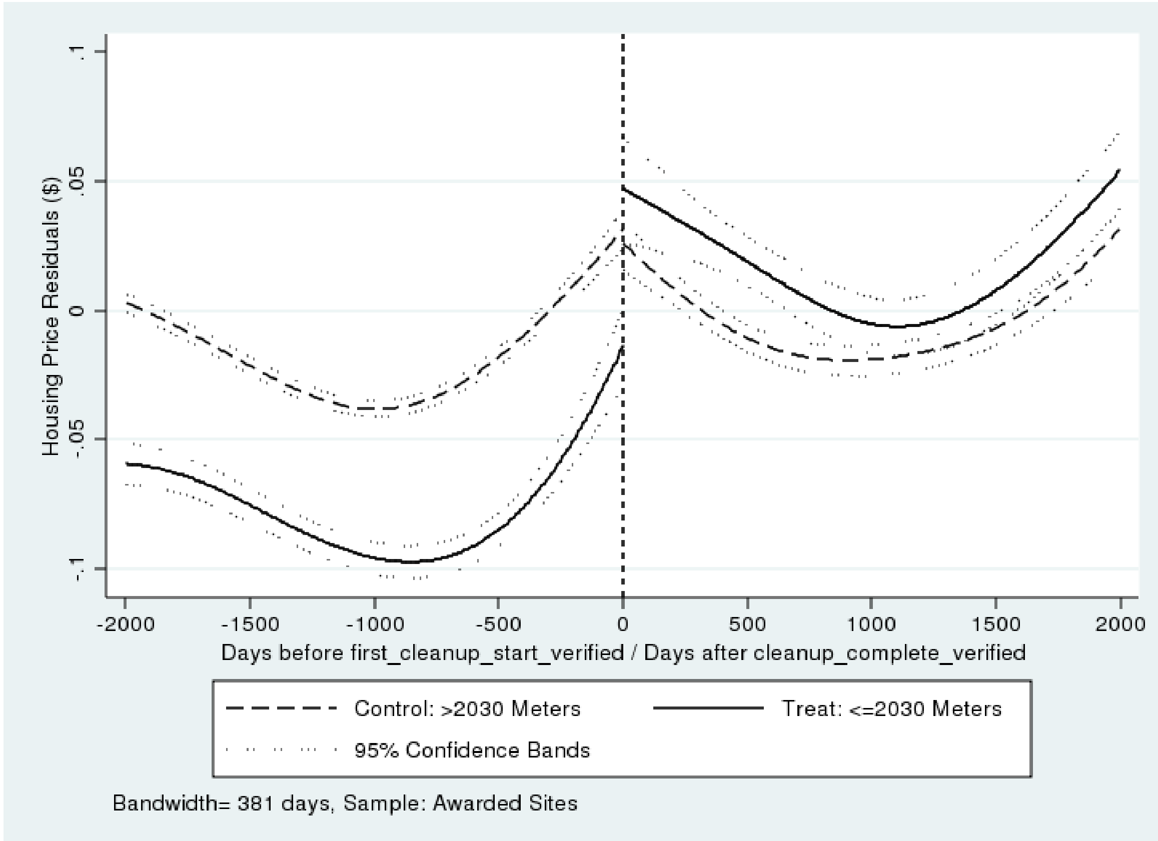


FIGURE 1.3: Non-Parametric Price Function Estimates Relative to Cleanup Period for Treatment and Control Houses With 95% Confidence Intervals

1.10 Appendix: Local Polynomial Modeling of the Hedonic Price Gradient

Let $(X_0^1, \ldots, X_0^j, \ldots, X_0^k)$ be a set of k equally-spaced focal points on the support of the variable defining distance from brownfield. Using k focal points divides the support of distance into $k+1$ intervals of length

$$l = \frac{dist_{max} - dist_{min}}{k+1}$$

where $X_0^j = dist_{min} + l \times j$ for $j = 1, 2, \ldots k$. We fit a linear function for each focal point:

$$P_i \mid X_0^j = a + b \cdot dist_i + \epsilon_i$$

where P_i is the price for house i and X_0^j is distance. The covariate and the focal points used in the kernel weight are normalized to have mean 0 and standard deviation 1. The problem is to minimize the following weighted sum of squared residuals,

$$\sum_{i=1}^{n} \left(P_i - [a + b(dist_i - X_0^j)]\right)^2 \cdot K_h\left(\frac{dist_i - X_0^j}{\hat{\sigma}}\right)$$

where $K_h(\cdot)$ is a Gaussian kernel; i.e. $K_h(z) = \frac{1}{h}K_h(\frac{z}{h}) = \frac{1}{h}\phi(\frac{z}{h})$, and $\hat{\sigma}$ is the estimated standard deviation of the covariate, X_i . The smoothing parameter h is chosen according to three times Silverman's Rule of Thumb, which states:

$$h = \frac{1.06\hat{\sigma}}{n^{1/5}}$$

Comparing the price gradients with respect to distance pre- and post- treatment, the estimates find that the difference becomes close to 0 at a distance from the brownfield of about 2 kilometers. Price gradients with respect to time are estimated similarly where the X variable is instead the days relative to cleanup initiation and completion.

2

Learning in a Hedonic Framework: Valuing Brownfield Remediation

2.1 Motivation

In the absence of prices for environmental amenities, hedonic property value models have been widely used as a revealed-preference approach to measuring the value individuals place on non-marketed goods. Rosen (1974) provided the theoretical foundations for interpreting the marginal equilibrium price of a non-marketed good as an individual's marginal willingness to pay (MWTP) for the good. Based on this model, many researchers have focused on recovering consistent estimates of the MWTP for amenities using changes in property values under quasi-experimental settings. Some examples include Chay and Greenstone (2005), Linden and Rockoff (2008), and Pope (2008). These studies exploit an exogenous change in a good (usually due to a policy shift) to establish a causal relationship between the amenity change and changes in local housing prices. The capitalization of the policy change into housing prices is then interpreted as the average MWTP for the amenity. Haninger et al. (2012) similarly exploits temporal and spatial variation in housing exposure to contaminated sites known as brownfields to estimate the capitalization of brownfield cleanup into local housing prices. What is crucial for these studies to recover unbiased estimates of the MWTP is that prices in the pre-policy period represent a valid baseline to which post-policy prices can be compared. However, this is likely to be complicated by the substantial provision of information to homeowners that accompanies policies such as brownfield remediation *before* the policy is implemented. Specifically, if consumers infer from information provided that cleanup of a nearby brownfield is likely (forward-looking behavior) or gain information about the severity of the brownfield contamination (learning from new information), then the baseline period prices need to be adjusted to consider these two effects as they may drive potential distortions to the MWTP estimate if left unaccounted for.

This paper considers distortions in policy valuation of non-marketed goods from not accounting for two types of household behaviors - an expectations bias from forward-looking behavior and an information bias due to learning from information provision. To address these two sources of bias, I incorporate learning into a dynamic hedonic framework. Using brownfield remediation as an application, I model a household's residential location decision as a dynamic, discrete neighborhood-choice problem, where households update their knowledge of brownfield hazard information in a Bayesian fashion before making decisions on where to live. Furthermore, I collect a new data set on brownfield hazard information, which allows me to test my model and estimate the bias. Since information provision is prevalent in policies under many settings, it is my hope to develop a general framework that can be applied to other areas in the future.[1] Section 2 describes these two information-related sources of bias in more detail, as well as other research that has addressed these biases. As part of the model is tailored to learning about a specific amenity, I next describe the

[1]Some examples of other amenities with public information provision include air quality, crime and school quality.

amenity of interest, a brownfield, in section 3. Section 4 then outlines a model of neighborhood choice that incorporates learning and forward-looking behavior. Section 5 gives summary statistics and discusses other data sources, followed by section 6, which describes the estimation procedure. Finally, I present the results in section 7, and conclude in section 8.

2.2 Two Sources of Information Bias

2.2.1 Expectation Bias

Rosen (1974) models households as static utility-maximizers. However, if individuals are forward-looking, their choices of amenities in the current period will reflect expectations for how these amenities may evolve, which will subsequently be reflected in transaction prices. This is most applicable in decisions relating to durable goods, such as that to purchase a house, given the size of the investment as well as the large costs associated with moving. Using a static framework to interpret MWTP from changes in housing prices will therefore confound the estimate with individual expectations about the future. This is exacerbated in the case of valuing the impacts of an amenity-related policy if there is sufficient information provision before the policy is implemented. Information provided may cause the policy of interest to be anticipated and drive an even greater expectations bias in the estimated MWTP.

In the environmental economics literature, studies have examined the role of information in the evaluation of environmental risk. Some examples include Brookshire et al. (1985), Hallstrom and Smith (2005), and McCluskey and Rausser (2003). These hedonic studies model the impact of information using a static, expected utility framework, where information changes affect individual choices through the probability of a certain (bad) event occurring. This is essentially a static approach to accounting for forward-looking behavior: households receive information about an amenity or event of interest, which in turn affects their perceptions of how the amenity will change, or the likelihood of the event occurring in the future. The recent urban and environmental economics literatures have seen dynamic structural models such as those of Bayer et al. (2011a), Bishop and Murphy (2011), Bishop (2012), and Mastromonaco (2011) that specify individual preferences and maximizing behavior to account for the forward-looking nature of household decisions. They employ the recent advances in models of dynamic discrete choice (Hotz and Miller (1993), Arcidiacono and Miller (2011)) to estimate the parameters of a household's MWTP curve for time-varying neighborhood amenities (e.g. crime and air pollution). The same logic applies to fixed amenities as well, since households may change their expectations of how likely a fixed (dis)amenity will exist in the future based on, among other things, information provided in the time leading up to the policy change.

2.2.2 Learning Bias

A second source of bias relates to information provision and learning from the information that is provided. Public disclosures may change what is known about an amenity. Many policies regarding the cleanup of a hazard will provide information to affected households before any remedial actions are taken. Furthermore, information regarding the amenities may be released in multiple waves. Suppose a household lives near an aesthetically displeasing site, and then learns that the site is not only unattractive but is also contaminated with low levels of carcinogens. It is likely that measured preference for the site will be different depending on whether baseline prices were measured before the information was released or after. Practically, this means that the estimated MWTP will depend on when one takes prices as baseline even if agents are assumed to be myopic. A solution could be to limit the use of baseline prices to those of houses sold after all relative information is released but before the policy is implemented. However, data in this narrow time frame may be sparse. Furthermore, as discussed earlier, the home-purchase decision encourages forward-looking behavior. Prices transacted right before a policy (e.g. brownfield cleanup) is implemented, may alleviate information bias, but could more heavily reflect the expectation that the nearby brownfield will be cleaned.

Previous work in environmental economics have recognized the importance of household knowledge of disamenities at the time of property transactions in the context of property value hedonics (Schulze et al. (1986), Michaels and Smith (1990), Gayer et al. (2000), Sanders (2013)). However, these papers have not explicitly modeled how information transmission may affect locational choices, which can be used to recover estimates from counterfactual scenarios that are important for policymaking. Other literatures provide examples of modeling learning's effect on individual choices (Ackerberg (2003), James (2011), Chan and Hamilton (2006), Chernew et al. (2008)). Although there are important differences in how information is acquired and processed between my paper and the latter group of papers, property value hedonic models are nonetheless derived from choices made by individuals, and thus learning from information released about nearby amenities should be considered. Whether the learning matters and is systematically priced into housing values is an empirical question, and depending on the answer, may affect the validity of MWTP estimates.

2.3 A Context for Brownfields

The disamenity of interest in my application is a brownfield. The U.S. Environmental Protection Agency (EPA) defines a brownfield as a 'real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant' (EPA (2012)). In 2002, the Small Business Liability Relief and Brownfields Revitalization Act was signed as an amendment to the Comprehensive Environmental Response, Compensation, and Liability Act of 1980. This created the EPA Brownfields program, which aimed to assist organizations in the redevelopment of brownfield lands through the provision of grants. The set of brownfield sites considered are those that submitted cleanup grant applications to this program from 2003 to 2008.[2] Data on brownfield sites originated from the EPA and were compiled by Kevin Haninger for the aforementioned work on brownfields (Haninger et al. (2012)). Crucial to the current paper, the data contain site information (exact location), applicant information, and dates relative to cleanup (if awarded).

The EPA brownfields data include detailed information regarding site remediation. However, in order to account for learning, the data need to be augmented with information updates about brownfield contamination. I infer the information available to households by collecting data from assessment reports, which document investigations into brownfield sites. The Massachusetts Department of Environmental Protection (MA-DEP) maintains an online database of contaminated sites, with uploads of original and/or scanned assessment documents. Each report contains a myriad of information regarding the history of investigative efforts and environmental condition of the site known *at the time* of the assessment (See Figures 1 and 2 for an example). Each site can have several associated assessment documents, ranging from the mid-1980s to the present. Consequently, for most sites, I can capture all investigations and associated results since the inception of the land as a contaminated site.

Implicit in the use of these assessments as a measure of general informed-ness is an assumption that households will retain information from investigations as the assessments are published. Since the database for contaminated sites is public information, households can retrieve information on sites by directly searching for them online, similar to what I have done. Furthermore, the Massachusetts Contingency Plan (MCP)[3] requires that legal notices for specific response action milestones be

[2]A drawback is that I do not observe information about brownfields that did not submit an application.

[3]The MCP is a set of official guidelines established by the state of Massachusetts regarding how

published in local newspapers in affected or potentially affected communities (For an example of a public notice, see Figure 3.). Given this, it seems plausible to assume that households can rely on these assessments to stay informed.

To construct the contamination data, I first match the brownfields in Massachusetts from the EPA data with those in the MA-DEP online database using address information. Next, I read through each assessment for information related to the investigation such as the date when assessment activities began, the date when the assessment was posted online, and the contaminants found to exceed state-defined thresholds in each exposure pathway (e.g. Polycyclic Aromatic Hydrocarbons (PAH) found in soil, as opposed to PAH found in groundwater). It is important to note that whether a site is designated as contaminated depends not only on the measured contamination levels at a site, but also on whether exposure risk was high. Specifically, the thresholds for contamination would be lower if there existed any institutions (schools) or environmental areas of concern (habitats) near the site. The thresholds are not available for all assessments, but it is important to account for them as they may affect the way in which households perceive released information. As a proxy for this, I use the Numerical Ranking System (NRS), which was developed by the MA-DEP to score a site according to the extent of a site's hazards and potential damage to its surroundings. According to MA-DEP regulations, a site must be classified using the NRS within a year of a toxin release notification (of Environmental Protection (2009)). The actual NRS score is constructed from the sum of five subscores, where subscores IV and V are respectively for the categories of Human Population and Land Uses, and Ecological Population. I retain subscores IV and V as a way of measuring the institutional and environmental areas of concern that may contribute to the assessment of the contamination. The NRS is only used when an initial release of hazard material is discovered, therefore I only observe one

to conduct investigations and cleanup for contaminated properties.

NRS score for most sites even though a site's surroundings may change.[4] This measure is imperfect. However, given the time frame of this paper, significant changes in institutions are likely limited, which leads me to believe that the NRS score is a reasonable proxy.[5] In addition, I record some historical information about the site including the year it became abandoned, and the type of business that had previously occupied the site. In total, I have located 65 sites in Massachusetts that map to the EPA data set, and have recorded information from between 1 and 5 assessments for each of these sites.

2.4 Model

Following previous work on dynamic locational choice (Bayer et al. (2011a), Bishop and Murphy (2011), Bishop (2012)), I model households making a sequence of residential location decisions in a finite-horizon framework. A simplification in my model as compared to that in Bayer et al. (2011a) is that I abstract from wealth accumulation and use moving costs as the driving factor for forward-looking behavior. Since the primary focus of this paper is to accommodate learning in a hedonic setting, this simplification seems reasonable given that it yields significant reductions to computational burden without compromising the paper's primary goal.

In each period, a household chooses whether to move, and conditional on moving, chooses the neighborhood that yields the highest value of expected lifetime utility. In practice, a neighborhood will be defined as a Census Tract. Household i's choice set at time t is described by $J+1$ alternatives, where alternative $J+1$ represents the choice to stay in its current residence.[6] In addition, if a household chooses to

[4] There are a few exceptions where scores are revised.

[5] Furthermore, a brief survey of the available thresholds in assessments shows that they very rarely change over time.

[6] If a household chooses a house within the same tract, it is considered a move and not a stay decision.

move, it incurs a moving cost in the same period.

The information structure of the model is such that households are uncertain about the transitions of neighborhood amenities in the future, but fully observe all neighborhood amenities in the current period with the exception of brownfield hazard. With regards to brownfield hazard, households only have imperfect information about a site's potential effect on health, and whether it will be remediated. In the face of this uncertainty over 'true' hazard levels, households can make inferences based on the results of periodic site assessments. Once their knowledge has been updated with the most recent assessment information, they can subsequently form a guess of how likely a particular site will be remediated after they have moved into a neighborhood.

The timing of the household's decision process is as follows, where the decision period is 2 years in length. At the beginning of time t, a household first makes appropriate updates to its beliefs about brownfield hazards if new information is available, and predicts whether sites will be remediated after they have moved in (if it has not already been cleaned). It then forms expectations over how other neighborhood amenities will change in future periods. Finally, after forming beliefs and expectations over attributes of neighborhoods that it has uncertainty over, the household chooses a place to locate for the rest of period t.

In the following subsections, I first detail the structure of the assessments, the learning process using the assessments, and then set up the household's problem.

2.4.1 Assessments

Environmental site assessments characterize the overall condition of a brownfield. The published results from these assessments serve as noisy signals about brownfield hazards. I quantify the assessment results into the sum of the total contaminants in each medium of contamination that exceed state toxin thresholds.[7] I assume that an assessment result for a site k in neighborhood j, c_{kt}, depends on the 'true', unobserved brownfield hazard level, denoted H_k, which does not vary over time. Households are also aware that these results additionally depend on the institutional and environmental settings around the site, IE_k. These IE settings determine the appropriate state thresholds with which to compare findings on toxin levels. As discussed in the previous section, different thresholds account for the use of more stringent standards when, for example, a school is located near a site. In absence of data on the thresholds, the NRS subscores for human and ecological populations will serve as proxies. Assessment results for a site k in neighborhood j are then given by

$$2.4.1\widetilde{c}_{kt} = H_k + \lambda \cdot IE_k + e_{kt} \tag{2.1}$$

where e_{kt} has an assumed i.i.d. normal distribution around 0 with variance σ_e, and λ is a parameter on the institutional and environmental settings around the site. Rather than directly using the count of the contamination, c_{kt}, in equation , I reparameterize the contamination as $\widetilde{c}_{kt} = \log(1 + c_{kt})$ in order to make hazard beliefs in neighborhoods with sites comparable to those without.[8] Since assessments do not occur in every period, a household at time t will refer to the information provided up to the most recent site assessment to form beliefs.

2.4.2 Learning

If a site has never undergone an assessment, households are assumed to have prior beliefs about hazards in the neighborhood that are distributed normally around H_0 with variance δ. When new assessment information becomes available, they compare

[7] For example, an assessment that finds lead contamination in soil and groundwater at a site yields a result of 2. In practice, I use 7 categories for contaminants, and 4 categories for the media of contamination. Categories are listed in Table 3 of the Appendix.

[8] Using the learning process detailed next in 4.2, signals are log normal so that only areas without sites can have a $\widetilde{c}_{kt}$ of 0, i.e. a hazard belief level that approaches negative infinity.

the new results with their prior beliefs and information received thus far to form updated beliefs. In particular, a signal for the hazard will be the difference between the new contamination results and the component attributable to the IE settings near the site,

$$sig_{kt} = \widetilde{c}_{kt} - \lambda \cdot IE_k = H_k + e_{kt} \tag{2.2}$$

The noise of each signal, in the context of contamination assessments, may result from variation in testing conditions or sampling methods at the time of the assessment. As households cannot separately disentangle the noise from the true hazard quality in each signal, the true quality of the brownfield can only be revealed after multiple signals. How quickly the true hazard is revealed depends on how noisy the signals are (in particular, in the case when signals have no noise, $\sigma_e = 0$, households can learn the true hazard level after one signal). Given the distributional assumptions on prior beliefs and the noise of the signals, household posterior beliefs will be normally distributed (DeGroot (1970)). The mean and variance of their hazard beliefs for a brownfield k in neighborhood j after receiving signals up to period t are given by

$$\begin{aligned} E_t(H_k) &= V_t(H_k)\left[V_{t-1}^{-1}(H_k) \cdot E_{t-1}(H_k) + \sigma_e^{-1} \cdot sig_{kt}\right] \\ V_t(H_k) &= \left[V_{t-1}^{-1}(H_k) + \sigma_e^{-1}\right]^{-1} \end{aligned} \tag{2.3}$$

These beliefs will serve as state variables that households factor into their neighborhood location decisions. Since there may be multiple brownfields within a neighborhood, households are assumed to learn about each brownfield separately, and then combine this information into beliefs for the overall hazard level for the neighborhood as $\{E_t H_j, V_t H_j\}$.[9]

[9]In this model, learning about a particular brownfield site occurs independently of other sites, although this can be relaxed.

2.4.3 Household Preferences

Household indirect flow utility from living in a neighborhood depends on the flow cost of housing, R_{jt}, the characteristics of the neighborhood j, and more generally the district in which the neighborhood is located, r_j. Neighborhood characteristics at t that are observable to the econometrician are summarized in X_{jt}, and include attributes such as crime, school quality, and poverty levels. Attributes that are unobserved by the econometrician, from which the household draws utility, are captured by ξ_{jt}. Also included in neighborhood attributes is the brownfield hazard, H_{jt}, which is neither observed by the econometrician or the household. This hazard only affects utility if brownfields in tract j are not remediated by the end of t, which the household believes to depend on the results of site assessments and other brownfield-specific characteristics. The final component of flow utility consists of costs to moving. A portion of this cost involves a financial moving cost that a household incurs if it chooses to move, MC_{it}. These are calculated as 6% of the value of the house a household moves from to represent the amount paid in realtor fees.[10] Another portion involves a psychological cost if it moves away from the district in which it was located last period. The last part depends on the distance it would have to travel to move to a specific neighborhood since farther moves can be more costly. This is denoted $dist(j, \cdot)$ for a move to neighborhood j. Let household i's choice of neighborhood at time t be denoted by d_{it}, and let the neighborhood variables (including hazard beliefs) and moving costs be summarized by s_{it}. Given the neighborhood characteristics and household beliefs over the true hazard, household i's expected utility of living in neighborhood j at time t takes the following linear

[10] Note that implies that conditional on moving, the household faces the same financial moving cost regardless of where it moves to. Values of what a house would have sold for if a household chooses to stay will be imputed.

form,

$$\begin{aligned} u_{ijt}(s_{it}) + \epsilon_{ijt} = {} & \beta_X X_{jt} + \beta_R R_{jt} + \beta_H E\Big(e^{H_j} \times \mathbf{1}[clean_{jt} = 0] \mid s_{it}\Big) + \xi_{jt} \\ & + \mathbf{1}[d_{it} \neq J+1] \cdot \Big(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}[r_j \neq r_{d_{it-1}}] + \beta_d dist(j, d_{it-1})\Big) \\ & + \epsilon_{ijt} \end{aligned} \tag{2.4}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function, and ϵ_{ijt} is an idiosyncratic shock to utility. Some comments are in order regarding the term relating to hazard. First, the hazard enters the utility exponentially to make hazard levels in neighborhoods with sites comparable to those without and those that have been cleaned. Second, there is evidence from cleanup grant application guides that the likelihood of award, aside from assessment results and site characteristics, may be higher for brownfields in localities that are either better at applying for grants (e.g. have more resources) or have more financial or socioeconomic needs. I will therefore allow cleanup to additionally depend on district- and tract- level characteristics. District-level variables include per capita spending and the unemployment rate. Tract-level characteristics include the percentage of people, in the year 2000, who are Black and Hispanic in the neighborhood where the brownfield is located. Site characteristics include the most recent contamination signal, the cumulative number of assessments performed, a dummy variable for groundwater contamination, and a dummy variable for whether a brownfield grant proposal was submitted. I summarize these other variables that the household perceives to affect cleanup award in Z_{jt}, which I will include in the set of state variables. Lastly, I assume that the beliefs for site hazard and cleanup are independent conditional on the state variables (which, most importantly, includes the contamination signals). This assumption requires the random component that determines whether a site is cleaned, which is not controlled for in s_{it} (e.g. an incomplete application, or a particularly low EPA budget in a given year) to be uncorrelated with site hazard. This allows me to re-write the expectation term in utility as $E\left(e^{H_j} \mid s_{it}\right) \times Pr(clean_{jt} = 0 \mid s_{it})$, so that the contamination will be scaled by the probability that brownfield sites in tract j will remain undeveloped. Lastly, in anticipation of needing to account for endogeneity issues related to unobserved neighborhood quality for estimation later on, I collapse all neighborhood- and time-specific variables in the flow utility into a mean utility term, θ_{jt},

$$u_{ijt}(s_{it}) + \epsilon_{ijt} = \theta_{jt} + \mathbf{1}[d_{it} \neq J+1]\Big(\beta_{MC} \cdot MC_{it} + \beta_{PMC} \cdot \mathbf{1}[r_j \neq r_{d_{it-1}}] + \beta_d dist(j, d_{it-1})\Big) + \epsilon_{ijt}$$

$$\text{where } \theta_{jt} = \beta_X X_{jt} + \beta_R R_{jt} + \beta_c E[e^{H_{jt}} \mid s_{it}] \times Pr(clean_{jt} = 0 \mid s_{jt}) + \xi_{jt} \quad (2.5)$$

2.4.4 The Household's Problem

The components of the flow utility make up the state of the world that is observable to a household at the time it makes its decision. I summarize the variables other than the idiosyncratic error in s_{it},

$$s_{it} = [X_t, R_t, E(H_t), V(H_t), MC_{it}, Z_t, \xi_t] \quad (2.6)$$

Given the above preferences, the household's problem is to choose a sequence of neighborhood locations (d_{it}) to maximize its expected sum of discounted flow utilities,

$$\max_{d_{it} \in \{1,\ldots,J+1\}} \mathbf{E}\left[\sum_{t'=t}^{T} \beta^{t'-t} u(s_{it'}, d_{it'}) + \epsilon_{it'} \mid \epsilon_{it}, s_{it}, d_{it}\right] \quad (2.7)$$

where β represents the discount rate for future consumption, and the expectation is taken over the idiosyncratic shocks as well as the transition of the state variables.[11]

Before proceeding, I make the following assumptions. (i) The idiosyncratic error terms are additively separable in the flow utility, which is already assumed given the setup of the flow utility. (ii) The idiosyncratic error terms are i.i.d. over households, alternatives, and time. (iii) The state variables, s_{it}, follow a Markov process, implying

[11]The discount factor is assumed to be 0.95^2 since the decision period is 2 years.

that the outcome at $t+1$ only depends on information at time t. (iv) Conditional on the choice and state, the idiosyncratic error term does not affect the expected outcome of the state next period. That is, writing the conditional pdf of s_{it+1} as $q(s_{it+1} \mid s_{it}, d_{it}, \epsilon_{it})$, this conditional independence assumption implies that

$$q(s_{it+1} \mid s_{it}, d_{it}, \epsilon_{it}) = q(s_{it+1} \mid s_{it}, d_{it})$$

Given these assumptions, I can write the household's problem recursively using a Bellman equation,

$$V_t\left(s_{it}, \epsilon_{it}\right) = \max_{d_{it}} \{v_t(s_{it}, d_{it}) + \epsilon_{it}\} \tag{2.8}$$

where

$$\begin{aligned} v_t(s_{it}, d_{it}) \ &\equiv \ u(s_{it}, d_{it}) \\ &+ \ \beta \int \sum_{s_{it+1}} V_{t+1}\left(s_{it+1}, \epsilon_{it+1}\right) q(s_{it+1} \mid s_{it}, d_{it}) dF(\epsilon_{it+1}) \end{aligned} \tag{2.9}$$

Here, I have defined a choice-specific value function, $v_t(s_{it}, d_{it})$, to be the value of choosing any neighborhood, d_{it}, plus the value from choosing optimally thereon after.[12] Equation 2.9 shows that since moving is costly and neighborhoods can potentially change, a household's location decision today can affect the values of the choices it faces in future periods. Furthermore, since there is uncertainty about how neighborhoods will transition, they can only guess at the future states and the associated utilities in expectation. For simplicity of notation, I will start writing $v_t(s_{it}, d_{it})$, the choice-specific value function for choosing d_{it} at time t, as $v_j(s_{it})$ (where the choice $d_{it} = j$).

[12]That is, the value function at $t+1$ is

$V_{t+1}(s_{it+1}, \epsilon_{it+1}) = \max\{v_{t+1}(s_{it+1}, d_{it+1} = 1) + \epsilon_{i1,t+1}, \cdots, v_{t+1}(s_{it+1}, d_{it+1} = J+1) + \epsilon_{iJ+1,t+1}\}$

Next, I follow much of the dynamic discrete choice literature in assuming that the choice-specific error, ϵ_{it}, is distributed Type I Extreme Value. This gives the following closed-form solution for the probability of choosing an alternative j,

$$P_j(s_{it}) = \frac{e^{v_j(s_{it})}}{\sum_{\ell=1}^{J+1} e^{v_\ell(s_{it})}} \tag{2.10}$$

Then, taking the distribution assumption on the error term together with Conditional Independence, the integrated value function also has a closed form (Rust (1987)), expressed as,

$$\int V_{t+1}(s_{it+1}, \epsilon_{it+1}) dF(\epsilon_{it+1}) = \log\left(\sum_{\ell=1}^{J+1} e^{v_\ell(s_{it+1})}\right) + \gamma \tag{2.11}$$

where γ denotes Euler's constant. The choice-specific value function can then be simplified as

$$v_j(s_{it}) = u_j(s_{it}) + \beta E\left[\log\left(\sum_{\ell=1}^{J+1} e^{v_\ell(s_{it+1})}\right) \mid s_{it}, d_{it} = j\right] + \beta\gamma \tag{2.12}$$

Conditional Choice Probabilities and Finite Dependence

I next follow Bishop (2012) in using the insights of Hotz and Miller (1993) and Arcidiacono and Miller (2011) to re-write the difference in the choice-specific value functions in 2.12 as a function of flow utilities and Conditional Choice Probabilities (CCPs), which yields my main estimating equation.

First, using Hotz and Miller (1993), the integrated value function in 2.11 can be mapped to functions of choice probabilities. In particular, I write the future value component of the value function using an arbitrary choice, k, and simplify the choice-specific value function in 2.12 as

$$v_j(s_{it}) = u_j(s_{it}) + \beta \sum_{s_{it+1}} \Big(v_k(s_{it+1}) - \log P_k(s_{it+1})\Big) \cdot q(s_{it+1} \mid s_{it}, d_{it} = j) \tag{2.13}$$

Next, given the setup of preferences, I employ finite dependence from Arcidiacono and Miller (2011) to limit the dependence of future states on previous choices. Along with the property of utilities that only differences matter, finite dependence allows me to remove the $v_k(s_{it+1})$ term on the RHS of 2.13, further simplifying household value functions. To be more specific, preferences are set up to have limited memory such that the utility a household derives from a choice at $t+1$ depends on where it is located at time t, but not where it is located before that. Since distance of a move is a state variable, a household's choice of where to live at t determines the moving distances to each of the other alternatives in its choice set at time $t+1$. In this way, the household's choice at t changes the state it faces at $t+1$. The choice at t, however, does not affect the state faced at $t+2$. This is since the distance between wherever a household chooses to locate at $t+1$ and locations chosen at $t+2$ will remain the same regardless of where it chose to live at t.[13] Now, consider two sequences of potential choices between t and $t+2$ for a household: $\{j,k,m\}$ and $\{g,k,m\}$. Under finite dependence, the expected value of choosing neighborhood m at $t+2$ conditional on choosing k at $t+1$ and j at t, is the same as choosing the same sequence of alternatives except with g as the first choice,

$$\sum_{s_{it+1}}\sum_{s_{it+2}} v_m\left(s_{it+2}\right) q\left(s_{it+2} \mid s_{it+1}, d_{it+1}=k\right) q\left(s_{it+1} \mid s_{it}, d_{it}=j\right) \tag{2.14}$$

$$= \sum_{s_{it+1}}\sum_{s_{it+2}} v_m\left(s_{it+2}\right) q\left(s_{it+2} \mid s_{it+1}, d_{it+1}=k\right) q\left(s_{it+1} \mid s_{it}, d_{it}=g\right)$$

Exploiting the fact that only differences in utility matter in a logit framework, that is,

$$P_j(s_{it}) = \frac{e^{v_j(s_{it})-v_g(s_{it})}}{1+\sum_{\ell\neq g} e^{v_\ell(s_{it})-v_g(s_{it})}}$$

[13]Household preferences rule out any form of 'experience' individuals may have accumulated in a particular place from the past, which makes using finite dependence possible in this case. Wealth accumulation would be problematic for using finite dependence. For a dynamic model that uses CCPs that includes wealth accumulation, see Bayer et al. (2011a)

and expanding $v_k(s_{it+1})$ in (14) with respect to choice m at $t+2$ (and similarly repeating this process for $v_g(s_{it})$), I can rewrite the probability of choosing an alternative as a function of flow utilities, 1-period-ahead choice probabilities, and transition probabilities alone, where the difference in choice-specific value functions is given by

$$\begin{aligned} v_j\left(s_{it}\right) - v_g\left(s_{it}\right) &= u_j\left(s_{it}\right) - u_g\left(s_{it}\right) \\ &+ \beta \sum_{s_{it+1}} \Big(u_k(s_{it+1}) - \log P_k(s_{it+1})\Big) q(s_{it+1} \mid s_{it}, d_{it} = j) \\ &- \beta \sum_{s_{it+1}} \Big(u_k(s_{it+1}) - \log P_k(s_{it+1})\Big) q(s_{it+1} \mid s_{it}, d_{it} = g) \end{aligned} \tag{2.15}$$

Since the value functions for choices j and g are expanded with respect to the same choice, k, at $t+1$, the neighborhood attributes in the one-period ahead flow utilities cancel, and so the value function difference simplifies to

$$\begin{aligned} v_j\left(s_{it}\right) - v_g\left(s_{it}\right) = u_j\left(s_{it}\right) - u_g\left(s_{it}\right) & \\ + \beta \Bigg(\beta_d \Big(dist(k,j) - dist(k,g) \Big) + \beta_{PMC} \Big(\mathbf{1}[r_j \neq r_k] - \mathbf{1}[r_g \neq r_k] \Big) & \\ + \beta_{MC} \Big(\mathbf{1}[j \neq J+1] \cdot MC_{j,t+1} - \mathbf{1}[g \neq J+1] \cdot MC_{g,t+1} \Big) \Bigg) & \\ + \beta \Bigg(\sum_{s_{i,t+1}} - \log P_k(s_{i,t+1}) q(s_{i,t+1} \mid s_{it}, d_{it} = j) & \\ - \sum_{s_{i,t+1}} - \log P_k(s_{it+1}) q(s_{it+1} \mid s_{it}, d_{it} = g) \Bigg) & \end{aligned}$$

The probabilities based on these value function differences can then be used to build the likelihood of household choices from which we recover the utility parameters.

2.5 Data, Summary Statistics, and Evidence of Learning

There are a total of 1361 census tracts in Massachusetts as defined in the 2000 Census. These census tract boundaries coincide with town/city boundaries, which will overlap with public school district boundaries for most towns. As previously noted, a neighborhood is defined at the tract level, and belongs to a town (which I will refer to as a district from hereon after). Since brownfield sites are fairly local disamenities, the use of tracts as neighborhoods aims to ensure that brownfields are capitalized into all houses in that neighborhood. With the use of a larger neighborhood definition, I may encounter an issue where a household chooses a neighborhood containing a brownfield, but the brownfield is located far away, thus making it difficult to measure the household's preferences for brownfields from its residential choices. I attach attributes (brownfields, crime, schooling, etc.) to each tract to describe the neighborhood using Graphical Information Systems (GIS) software (See Figure 4 for a map of brownfields to census tracts).

2.5.1 Brownfield Summary Statistics

For the 65 brownfield sites in my sample, a total of 223 assessments were performed between 1984 to 2012. Figure 5 gives the number of assessments performed over time. Each site had between 1 and 5 assessments, where the average and median number of assessments were around 3 for each site (Table 2.1).[14] Although the magnitudes of contamination are small on average compared to the level, we can reject the hypothesis that the average change in brownfield contamination is equal to 0 against the alternative that it is greater than 0, which suggests that sites were found to be slightly worse with each assessment ($\Delta c_{jt} = c_{jt} - c_{jt-1} > 0$). However, there were instances when contamination was reported to improve ($\Delta c_{jt} < 0$), or not

[14] Only 1 site had 1 assessment.

change at all ($\Delta c_{jt} = 0$). Although I have assumed that true contamination does not vary over time, this is not contradictory to that assumption, as it is possible that technological increases have allowed for more comprehensive testing over time.

For sites with more than one assessment, the average time interval between the first and last assessment was 4.5 years, with a standard deviation of 3.76 years. Figure 6 plots the assessment intervals for all sites. The model from the previous section takes the arrival of these assessments as being exogenous. Given the range in assessment interval, and the variation in the number of assessments performed across sites, it seems reasonable to believe that households do not form expectations for the arrival of assessments (i.e. they are a surprise). From surveying assessment reports, most site investigations are initiated due to a report being filed with the MA-DEP that noticed a change in the site (e.g. foul odor, leaks). A concern might arise if sites in wealthier neighborhoods are assessed more often because the residents in these neighborhoods are more likely to report. To check for this possibility, I construct two measures of assessment frequency, the number of assessments per year and the number of years since the last assessment, and regress these measures on attributes of the neighborhood in which sites are located (Table 12). Although parameters on these attributes are statistically significant, the magnitudes of these parameters and of the average changes in the independent variables seem to imply that they are not economically significant. Although not a definitive test for exogeneity, this alleviates some concerns that related unobserved variables might be driving the arrival of information.

In terms of site type, the majority fall into the categories of either Manufacturing (e.g. paper, clothing, boilers), Other Services (e.g. auto repair, parking), or Trade, Transport, and Utilities (e.g. warehouses, storage facilities, gas stations) as shown in Table 2. The largest contaminant category was petroleum products, followed by contamination from Volatile Organic Compounds (VOCs), Polycyclic Aromatic

Hydrocarbons (PAHs), and metals (Table 3). Many of these contaminants are a result of the chemicals used in the process of manufacturing products for businesses that previously occupied the site.

2.5.2 Other Neighborhood Amenities

Neighborhood rental prices are imputed from median neighborhood transaction values according to Himmelberg et al. (2005) and Poterba et al. (1991).[15]

Data describing exposure to crime are taken from the FBI Uniform Crime Reports. This gives the number of violent crimes as reported by police agencies (corresponding to towns/cities) in Massachusetts. To get crime in per capita terms, I divide by the population for each district over time. The crime data is only available at the town/city level up through 2011. As a measure of school quality, I use the percentage of grade 10 students in each school district that achieved a score of Advanced or Proficient in math on the Massachusetts Comprehensive Assessment System (MCAS) test. For a measure of poverty, I use the percentage of students in each district that come from low income households. Both measures are available at the school district level from 1998 to 2011 from the Massachusetts Department of Education website.

Massachusetts district level attributes used to predict the probability of cleanup

[15] Annual cost of ownership is calculated as

$$R_{jt} = P_{jt} \cdot \left[r_t^{risk\ free} + r_t^{property} - \tau_t^{inc} \left(r_t^{mortgage} + r_t^{property} \right) + \delta_t - g_{t+1}^{cap\ gain} + \gamma_t \right]$$

The risk free interest rate, $r_t^{risk\ free}$, is the rate on the 3-month U.S. treasury bond. The property tax rate, $r_t^{property}$, is given annually by city/town, and the state income tax, τ_t^{inc}, is given annually. Both are taken from the Massachusetts Department of Revenue website. The mortgage rate, $r_t^{mortgage}$, is taken from FHFA annual mortgage rates for single-family homes in the Boston MSA. Expected capital gains, g_{t+1}, are taken to be the sum of expected inflation and real appreciation in housing price. Expected inflation is taken from the Survey of Professional Forecasters (SPF) maintained by the Philadelphia Federal Reserve Bank, which gives quarterly inflation forecasts. Expected real appreciation in housing prices is taken to be the spread between long- and short- run interest rates. The long-run interest rate forecast used is the rate on the 10-year U.S. treasury bond is taken from the Livingston Survey. Depreciation is assumed to be 2.5%, and the risk premium for owning versus renting is assumed to be 2% (Flavin and Yamashita (2002)).

are more generally taken from www.mass.gov, the official website for the Commonwealth of Massachusetts. These include district-level indicators that vary over time such as unemployment, budgeting, and population. Race variables are taken from the 2000 Census. Summary statistics are provided in panel A and B of Table 4.

2.5.3 Housing Data

The real estate data are from Dataquick Information Services. The time frame I consider is 1998 through 2011 for the state of Massachusetts. The data contain the exact location and characteristics of the universe of housing transactions. For each house in the data set, which is tracked by a unique property identifier, house attributes, such as square footage, age, and the number of bedrooms and bathrooms, are recorded from the most recent tax assessment. Information about the history of transactions for each property includes the transaction price and date, the names of the buyers and sellers, as well as information about the buyer's mortgage loan.

Although Dataquick provides unique property identifiers, it does not provide identifiers for the households that move into and out of the properties. Using buyer and seller names, and the transaction dates, I obtain information on a household's last location decision. I accomplish this by following a name-matching algorithm used in Bayer et al. (2011a). The algorithm takes the first and last name of a buyer, and looks for a seller with the same first and last name within a window of a year of the transaction. Upon matching where a buying household came from, I check to see whether the specific house sells again based on the unique property identifier. If it does, then I infer that this household chose to stay in their current location at each period up until the date the house transacted again, though I do not track where it then moves afterwards. In the resulting data set, for each move decision (an observed sale transaction), I know where the household is moving from, and for each stay decision, I know where it is choosing to stay.

The analysis limits transaction types to arms length transactions and properties that are owner-occupied. Since only housing characteristics from the most recent assessment are recorded, houses indicated to have undergone a major improvement after the beginning of my sample time frame are dropped, as attributes would be incorrect for previous transactions otherwise. Houses that sold more than once per year or four times per the window of house sales (14 years) are excluded. Prices are normalized to January 2000 dollars using the monthly All Urban Consumer Price Index for Housing in the U.S. Northeast Region available from the Index (2012). The analysis then excludes the 1st and 99th percentile of the observed price distribution.

After cleaning the housing data, I observe sales in 1300 tracts across 14 years between 1998 and 2011. Of 953,509 transactions after cleaning the data, I identified previous locations for 158,319 transactions (approximately 17%). This is a clear drawback of using a name matching algorithm, as I will not be able to track renters who become homeowners, those who move from outside of Massachusetts, and those who change names. There is also a concern that prior locations for buyers with more common names could be mismatched. These are problems suffered by all analyses of housing market dynamics, and highlight the need for collection of better data on residential mobility of homeowners. The final sample includes 158,319 choices to move, and 520,251 choices to stay. Panel C of Table 4 summarizes the house attributes for these transactions.

2.5.4 Evidence from the Data

The learning parameters are identified by the way prices respond to information signals. Thus, before proceeding to estimation, it is worthwhile to discuss identification and look for evidence of learning in the new data. Assuming that households are Bayesian learners implies that they learn about the unobserved brownfield hazard in a very specific way (i.e. using recursive Bayesian updating formulas), and that this way of learning is optimal given the distributional assumptions on the noise and prior beliefs. The means of posterior beliefs on hazard are determined by whether prices after information is released are, on average, different compared to those before signals are released. I allow the prior mean to be determined in a similar manner. Most papers that use Bayesian learners assume, taken to my context, that the prior mean for brownfield hazards should be the average over all posterior means. That is, on average, agents are rational and have correct beliefs. The learning setup in my paper departs from this assumption because if households are on average correct, then learning should not matter in the average MTWP estimate. Assuming informed contamination priors restricts how much could be learned about brownfield hazard.[16] The variance in signal and prior noise is identified based on the number of information signals it takes for prices to become stable (i.e. when signals stop imparting information). The parameters on IE settings are determined by how perceived information signals affect household beliefs about hazard.[17]

To see whether learning is translated to the data, I first look at the average impact of the first assessment for each site. Figure 7 gives a nonparameteric plot of housing price residuals against the days relative to the first assessment, where the vertical line denotes the day when the first assessment is performed (prices are adjusted for house and neighborhood attributes as well as year and district fixed effects). In an attempt to isolate the effect of that assessment, I limit the observations to a window of 2 years before and after the initial assessment date. There is a discrete (and statistically significant) downward jump in equilibrium prices at the first assessment date. If I restrict my model by assuming homebuyers have high priors relative to

[16]In fact, when estimating my model, which will be discussed in the next section, with a prior mean set to be the average of the posterior means, the difference in the estimates with and without learning disappears.

[17]If households know that more institutions nearby implies the standards for a contamination designation are lower, then households should interpret the signal for hazard to be relatively low given high IE settings compared to a place with the same signal and low IE settings.

the first assessment (i.e. they initially believe these areas to be fairly contaminated), then my model would not be able to explain this fall in prices after initial assessments are released. I repeat the exercise in Figure 7 with regressions in Table 13 (Figure 7 would correspond to the regression in the first column - 'First Assmt'). The estimates imply that on average, controlling for house attributes, neighborhood characteristics, and year and district fixed effects, assessments are imparting new information to the market. Other than the last assessment before cleanup, prices fall, on average, after assessments are released. Consequently, the changes in prices after each assessment suggest that learning occurs over the course of information releases.

2.6 Estimation

Estimation will proceed in four stages. First, given a guess of the prior mean, H_0, I estimate the other learning parameters from the contamination signal equation to form beliefs about brownfield hazards. I also estimate parameters that determine cleanup in this step. Second, I estimate the household's dynamic neighborhood choice problem to recover the moving cost parameters and a mean utility for living in each neighborhood at each time period. In the third stage, I decompose the mean utility estimates with respect to neighborhood attributes, including the hazard beliefs. In the final stage, I perform a grid search on the prior mean, repeating steps 1 and 3 for each guess of the prior, and retain the learning and the utility parameters associated with the highest likelihood from the mean utility decomposition.

Estimating the problem in this order boils down to the following assumptions. First, whether a site is cleaned does not depend on where people choose to live. Second, households only learn about brownfield hazards through the released contamination signals. Finally, cleanup only depends on hazard through the contamination signals. To see how, we observe from the data whether a person moves, and the conditions in the place they choose. The contribution to the conditional likelihood for household i at t given the state variables, s_{it}, is characterized by the joint pdf of the observed location decision and the contamination results for all neighborhoods up to time t. Writing the joint density as the conditional times the marginal, the likelihood can be split as done in the following

$$\begin{aligned} &\ell(d_{it}, c_1 \dots c_t, clean_{jt} \mid s_{it}, H) \\ &\quad = \ell(clean_{jt} \mid d_{it}, c_t, s_{it}, H) \times \ell(d_{it} \mid \{c_k\}_{k=1}^{t}, s_{it}, H) \times \ell(\{c_k\}_{k=1}^{t} \mid s_{it}, H) \end{aligned} \tag{2.16}$$

Assuming that household choices do not affect cleanup, cleanup's dependence on household choices can be removed. The likelihood can also be re-written to depend on the hazard using the law of total probability,

$$\begin{aligned} &= \ell(clean_{jt} \mid c_t, s_{it}, H) \times \ell(d_{it} \mid \{c_k\}_{k=1}^{t}, s_{it}, H) \times \ell(\{c_k\}_{k=1}^{t} \mid s_{it}, H) \\ &= \int_H \ell(clean_{jt} \mid c_t, s_{it}, H) \ell(d_{it} \mid \{c_k\}_{k=1}^{t}, s_{it}, H) \times \ell(\{c_k\}_{k=1}^{t} \mid s_{it}, H) dF(H) \end{aligned} \tag{2.17}$$

Next, as James (2011) noticed in a different context of occupational ability learning using wages and academic grades as signals, if we assume that household choices and site cleanup only depend on hazards through the observed sequence of contamination outcomes, then conditioning on the hazard in addition to the contamination results provides no additional information for observing household choices or whether a site will be cleaned. In other words, the contribution to the likelihood can be written

$$\begin{aligned} &= \int_H \ell(clean_{jt} \mid c_t, s_{it}) \times \ell(d_{it} \mid \{c_k\}_{k=1}^{t}, s_{it}) \times \ell(\{c_k\}_{k=1}^{t} \mid s_{it}, H) dF(H) \\ &= \ell(clean_{jt} \mid c_t, s_{it}) \times \ell(d_{it} \mid \{c_k\}_{k=1}^{t}, s_{it}) \times \int_H \ell(\{c_k\}_{k=1}^{t} \mid s_{it}, H) dF(H) \end{aligned} \tag{2.18}$$

Rewriting the likelihood in this way suggests that estimation of the hazard beliefs and the utility parameters that generate household choices can proceed sequentially. Specifically, I can solve the household's discrete choice problem to recover utility parameters, taking the estimated hazard beliefs as a state variable. Furthermore, recall that (1) hazard beliefs are formed at the level of the neighborhood, and (2) I have collapsed all neighborhood- and time- level variables into a mean utility term for the discrete choice estimation. In this case, if the neighborhood state variables can be summarized by the mean utility terms and households forecast the transitions of these mean utilities to determine their choice-specific future values (rather than forecast each component of the state separately), then I do not need hazard belief estimates until after I have recovered these mean utilities. This implies that I can circumvent re-estimating the discrete choice problem to evaluate another guess for the prior mean, since the prior (and hazard beliefs) would only affect how the mean utilities decompose into different neighborhood attributes. In the following sections, I outline each stage of the estimation in detail. In the final part of this section, I also discuss a measure of the value of information that can be constructed with the estimated parameters.

2.6.1 Stage 1: Posterior Beliefs

In the first stage, I estimate the linear model of contamination results in equation 2.4.1 on site hazard and IE surroundings given a guess of the prior mean (H_0). The goal of this stage is to recover and use the parameters of the contamination model to characterize household beliefs from different exposures to site information at each point in time. Although a site's IE surroundings are observed data, I do not have data on true site hazard as it is unobserved. To deal with this, I follow James (2011) in using an Expectation and Maximization (EM) algorithm (Dempster et al. (1977)). The EM algorithm provides a way to estimate the learning parameters, based on unobserved data (H). Recall that contamination signals for a brownfield k in neighborhood j are given by

$$sig_{kt} = \widetilde{c}_{kt} - \lambda \cdot IE_k = H_k + e_{kt}$$

where e_{kt} is distributed $N(0, \sigma_e)$. If a site has never experienced any signals, the initial prior on the hazard level is distributed $N(H_0, \delta)$. Thus, the learning parameters we want to estimate, excluding the prior mean, are

$$[\lambda, \sigma_e, \delta]$$

The EM Algorithm is an iterative procedure that estimates these parameters by imputing the unobserved data given a parameter guess, and then using the data to build the likelihood function that is then maximized to recover a new guess of parameters. I describe the two steps in detail in the Appendix. For neighborhoods that do not contain brownfields, I assume the belief is that there is no contamination with certainty.

2.6.2 Stage 2: Dynamic Discrete Choice

Stage 2 recovers the moving cost parameters $(\beta_d, \beta_{MC}, \beta_{PSY})$, and the mean utilities, θ_{jt}, for each neighborhood at every time period using household location decisions, leaving the parameters on specific neighborhood attributes to be estimated in a later stage.

Recall the value function difference, $v_j(s_{it}) - v_g(s_{it})$ derived in the model section. The future value component of this difference can be approximated by one-period ahead flow utilities and Conditional Choice Probabilities (CCP) of some alternative k at $t+1$,

$$\beta \cdot \Big(u_k(s_{it+1}) - E\left[\log P_k(s_{i,t+1}) \mid s_{it}, d_{it} = j\right]\Big)$$
$$- \beta \cdot \Big(u_k(s_{it+1}) - E\left[\log P_k(s_{i,t+1}) \mid s_{it}, d_{it} = g\right]\Big)$$

It follows that to build the likelihood, we need to first forecast values of the state variables at $t+1$ based on the state at t, and then predict the one-period ahead CCP's based on the forecasted state next period. Since I have collapsed all neighborhood-time level terms into θ_{jt}, I will forecast θ_{jt+1} rather than predict state variables at

$t+1$ individually. To do this, I first estimate moving cost parameters and mean utilities using a reduced-form multinomial logit.[18] Assuming that households believe the mean utilities follow an AR(1) process, I then pool the estimated θ_{jt}'s across neighborhood and time, and regress the current value of the mean utility on its previous value and dummy variables for each time period. I then combine draws from the residual distribution estimated from the AR(1) process with the current values of θ_{jt} to predict potential values in the next period.[19] Given the reduced-form logit parameters and the draws of the state at $t+1$, I can form the probability for choosing alternative k for each draw of the predicted mean utility. Averaging over the calculated probabilities from each draw then gives the CCP for choice k in period $t+1$.

Once I have recovered approximations of the future value component of household value functions, I can use them to build and maximize the following likelihood,

$$\ell\left(\beta_d, \beta_{MC}, \beta_{PSY}, \theta\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J+1} \mathbf{1}[d_{it}=j] \cdot \log\left(\frac{\exp\left(v_j(s_{it}) - v_g(s_{it})\right)}{1+\sum_{m\neq j}\exp\left(v_m(s_{it}) - v_g(s_{it})\right)}\right)$$

Since the value function is linear in the remaining utility parameters, this is simply a multinomial logit with an adjustment term for forward looking behavior, where the mean utilities are recovered using a Berry (1994) contraction mapping at each iteration of calculating the likelihood.

2.6.3 Stage 3: Preferences for Neighborhood Attributes

In the third stage, I recover preferences for neighborhood level attributes by decomposing the estimated mean utilities in equation 2.5. Concern for estimating that

[18]This is done using a Berry (1994) contraction mapping, where the mean utilities within each time period are normalized with respect to one choice. In practice, to speed up estimation, I estimate a 3-level nested logit and constrain the nesting parameters to be 1 at each level. The nesting of the decision process starts with the choice to move or stay, followed by the choice of district given the choice to move, and then the choice of tract conditional on the choice of district. Parameters for the nested logit are given in Table 14 of the Appendix.

[19]In practice, I take 25 draws from the residual distribution.

equation directly arises because of the endogeneity of price, as the costs of living in a neighborhood may be correlated with unobserved factors about the neighborhood. Assuming that the marginal utility of a dollar in moving costs is the same as that for housing costs, I can use the moving cost parameter estimated from the discrete choice step to directly control for the effect of housing costs on mean utility, $\widehat{\theta}_{jt} - \widehat{\beta}_{MC} R_{jt}$. I then regress these adjusted mean utilities on neighborhood attributes to recover the MWTP, and control other neighborhood unobservables with year and district fixed effects.

2.6.4 Stage 4: Estimate the Prior Mean

In the final stage, I revise the guess of the prior mean, re-estimate the learning parameters (stage 1) based on the revised prior, and get MWTP estimates (stage 3) based on this updated guess. Guesses are made along a grid from -6 to 6 at an interval of 0.05. The likelihood from the mean utility decomposition will be used as the criteria for choosing the prior mean estimate.

2.6.5 Value of Information

In the process of better estimating household MWTP for site remediation under learning, we can see whether households' decisions are impacted by what they learn. If learning does in fact occur, then it would suggest that assessments are a valuable tool for more informed decision-making. Once the utility parameters are recovered, we can calculate the value of the information provided by the site assessments. The DeGroot (1970) measure of the value of information for a person i at time t is defined as the difference in the utility achieved from the optimal choices under pre- and post-information sets,

$$V_{it}^{I} = \sum_{i} \left[V_i \left(\tau_1, d_{it}(\tau_1) \right) - V_i \left(\tau_1, d_{it}(\tau_0) \right) \right]$$

where τ_t is information at time t, and $d_i(\tau_t)$ is the optimal choice under the information set at time t for household i. Since in this setting, information is continually released over time, I set the 'post' information set to be the state of information resulting from the assessments that actually occurred (τ_1), and the 'pre' information set to be that if the last assessment for a particular site before cleanup was never performed (τ_0). Intuitively, V_{it}^I is the difference between the utility from what a person actually chose and the utility achieved from making a sub-optimal decision that would have been optimal under the state of knowledge associated with one less assessment being performed for that site, e.g. the value of the marginal assessment.

2.7 Results

I estimate the prior mean on hazard, H_0, to be -1.4 (Figure 8 plots the log-likelihood of the grid search). Compared to an average for the most updated posterior mean of $\overline{E_T(H_k)} = -0.48$, household initial beliefs about hazard were, on average, 1 standard deviation lower than what was revealed later on. Stage 1 results for the learning parameters are given in Table 5. These parameters are used to build the posterior distributions on hazard for each neighborhood at each time period. As an example, I plot the belief distribution for one of the brownfield sites in Figure 9. As expected in Bayesian updating, the mean of the posterior beliefs fluctuates in the direction of the contamination signal, and the uncertainty over the hazard distribution is reduced with each signal that is received.

Stage 2 consists of first estimating the Conditional Choice Probabilities (CCPs) to get the future value portion of each household's value function, followed by using the CCPs in the estimation of the household's dynamic discrete choice problem. Using the CCP estimates, Table 6 gives the discrete choice estimates for various parameters associated with moving costs, which are the remaining parameters not absorbed into the neighborhood-time mean utility term.[20] I find physical moving costs to be \$105.24 per kilometer. Given that the average distance of moves is 48.82 kilometers, this implies an average physical moving cost of \$5,137.90.[21] I include estimates for the myopic case, where the future utility portion of the value function is set to 0 (in other words, the discount factor, β, is 0).

With the estimates of neighborhood mean utilities, stage 3 decomposes the tract-time mean utilities to recover MWTP for the hazard, as well as other attributes. In order to evaluate the effect of modeling learning and forward-looking behavior on MWTP estimates, it is useful to compare the estimates for the model outlined from the previous sections with an estimate that either (i) assumes myopic households, (ii) assumes full information about brownfield hazards so that learning is not needed, or (iii) does both (i) and (ii). As previously mentioned, adding dynamics in my context implies allowing households to be forward-looking with respect to time-varying amenities ($\beta \neq 0$), and also allowing households to forecast the cleanup likelihood of nearby brownfields in the future.[22] Modeling households as static utility maximizers when they are forward-looking may lead one to observe higher levels of housing values with, for example, high levels of crime. From this, we would conclude that people care little for crime, when, in reality, they are simply willing to pay higher prices today because they expect this disamenity will improve in the future. Assuming full information on contamination, which I will refer to as the 'no learning' case, means that we assume households are fully informed about each site's contaminants even

[20] Table 14 provides the estimates used to construct the CCPs.

[21] For a 3.4 bedroom apartment, the mean in my sample, Bieri et al. (2012) would predict physical moving costs to be between \$2,500 and \$3,500. The moving cost estimates in Bieri et al. (2012) are based on converting average bedroom size to a weight in pounds that an average household would have to transport for a move, and then calculating the cost of transport for varying travel distances at various weight groups. Although my estimate is larger than what they predict, their estimate is for renters whereas my estimates are based on a sample of homeowners, whom may presumably have larger physical moving costs.

[22] Table 15 gives estimates for the probability of cleanup.

when they may actually be uninformed. This raises a concern that we might then observe high levels of housing values associated with high levels of contamination simply because households do not know the contamination exists, which can then lead to a downward bias in the measured MWTP. I can use my model to recover an estimate of the MWTP assuming no learning by simply altering the contamination exposure for all households to be the contamination reported from the last assessment for each brownfield (before it is remediated). This will replicate the information structure in the standard hedonic framework. Modeling learning can then be thought of as adjusting the contamination levels to reflect households beliefs at a given point in time.

Table 7 presents the raw estimates from the mean utility decomposition for each of the four cases. Panels A and B respectively give the estimates assuming myopic and forward-looking households. Within each panel for both the learning and no learning cases, utility decompositions are shown with and without year and school district fixed effects. Standard errors that adjust for the estimation error in this multi-stage estimation procedure will be bootstrapped, however, first stage standard errors are temporarily provided for reference. Focusing on the columns that include fixed effects, parameter signs are as expected for hazard, per capita crime, % low income, and average house attributes. Although the school quality measure, which is the % of students with an Advanced or Proficient MCAS score, shows the opposite sign from what one would expect, the variable is insignificant even with first stage standard errors. This result is likely to occur with school district fixed effects when the areas with high levels of school quality are always the same districts. The inclusion of these neighborhood attributes mainly serve as controls that proxy for other unobserved characteristics about neighborhoods that might drive sorting behavior beyond what fixed effects can control. Although signs are generally sensible, caution must be exercised in interpreting the magnitudes as they may be biased by correlations with unobservables.

Table 8 converts the estimates (with fixed effects) into dollar values using the financial moving cost parameter. Comparing across columns, Panel A of Table 8 shows that the MWTP to avoid a rise in a unit of contamination in the myopic and forward-looking cases are about the same when learning is modeled ($891.23 compared to $888.38). This difference is slightly larger without learning ($363.81 compared to $416.52, although the confidence intervals, based on OLS standard errors, overlap). The overall similarity between the myopic and forward-looking estimates suggests that households are *not* anticipating sites to be cleaned upon receiving contamination signals. Furthermore, the bias in not allowing for learning is alleviated when forward-looking behavior is allowed, as evident from a larger MWTP estimate when dynamics are included. However, the interpretation of the source of this bias is quite different. If one only examines the estimates without learning, they may incorrectly attribute bias to households being forward-looking as opposed to having incomplete information neighborhood attributes.

In terms of learning versus full information (comparing across rows in Table 8, Panel A), the MWTP with learning in the dynamic case is more than double the estimate without learning, rising from $416.52 per unit of contamination, to $888.38. The bias is even larger when forward-looking behavior is removed from the models. This implies that households are slowly learning about contamination over time, and in particular, find that contamination is worse than what they had originally thought. Thus, when we take data on housing values to infer MWTP and assume households are informed about site hazards, we are falsely attributing high housing values to households not caring (enough) about brownfields when in fact the households are just uninformed about the extent of the contamination.

It is also useful to compare these estimates to the simple hedonic model. Table 9 gives results from a hedonic regression of rental price regressed on contamination levels, where the contamination is assumed to be from the last recorded assessment before cleanup. For consistency, I include the same set of controls used in the mean utility decomposition. The hedonic model estimates a MWTP of \$373.30 to reduce a unit of contamination, which is consistent with the static estimate without learning in my model (\$363.81).[23] Given that the two models seem comparable, I compare the hedonic regression estimate to the estimate that accounts for learning and forward-looking behavior and find that the MWTP estimate in my model is 2.38 times the simple hedonic estimate.

As brownfield sites are fairly diverse, one might expect willingness to pay to vary depending on what sites will be converted to after cleanup. Table 10 additionally adds contamination interacted with dummy variables for future use in the forward-looking model with learning. The set of variables with which I interact contamination include (1) a dummy for whether future use is known, (2) a dummy for if a site will be used for green space, and (3) a dummy for if a site will be used for parking. The base group to interpret the interaction terms would be the group for which future use has yet to be determined. I find that if future use is unknown, MWTP for a unit of contamination is only \$162.01, compared to \$812.61 (=\$162.01 + \$650.59) if future use is known. Furthermore, if the site is used for green space, the MWTP increases to \$1,023.08 (=\$162.01 + \$650.59 + \$210.47). Site conversion to a parking lot adds \$107.30 to MWTP for contamination at a site with known future use, although this is statistically insignificant under first stage standard errors.

Finally, using the parameter estimates, I can calculate the DeGroot measure of the value of information, which gives the average utility loss in dollar terms of removing one assessment. I calculate this for every site in my sample and present an average. Since a change for a particular site will most directly impact households in

[23]Since the simple hedonic estimate does not include moving costs, the model for the myopic/no learning case is still slightly different from the model that produces the hedonic estimate. Thus, we should still expect some differences in the estimates.

the tracts near the site, I calculate this for the sample of households who actually chose neighborhoods that are within 3km of sites. The average of the loss in lifetime utility per household is \$17,887.38, or a flow value of \$894.37 (Table 11, Panel A). I separately calculate the value of information for sites by previous use, categorized by the NAICS codes (Table 11, Panel B). The value of information ranges from \$14,089.69 to \$23,307.93, with the lowest value belonging to recycling and waste facilities (categorized as Professional and Business Services), and the highest belonging to those used for financial activities (office space). I additionally calculate the information value for some specific types of sites (Table 11, Panel C). The value of information from an assessment of a former gas station (\$16,240.80) is relatively small compared to that from an assessment of a former factory (\$19,209.43) or a school, park, or office building (\$22,771.54). This ranking is intuitive as there is more value in revealing contamination for locations that household do not expect to be polluted.

2.8 Conclusion

This paper builds learning and expectations into the traditional hedonic framework as a way of improving upon existing methods to value environmental amenities. Specifically, the aim is to correct an information-related bias to the MWTP estimate, which, in the context of brownfield cleanup policy, translates into a faulty assumption that households have full information about a site's contamination levels. To identify whether this bias exists and, if so, its direction and magnitude, I collect data from assessments on brownfield contamination over time for 65 sites in Massachusetts. The published assessments are assumed to convey information about brownfield hazards to households. In addition, since dynamics have been shown by recent work to be a source of bias to amenity valuation, the information-related bias cannot be simply recovered by estimating the hedonic gradient while separating out the housing values associated with different information sets. Instead, one should allow for the possibility that the released information may impact household expectations for how a site might evolve in the future. Thus, in order to account for both learning and forward-looking behavior, I build upon previous dynamic neighborhood choice models by additionally allowing households to learn about brownfield contamination while making residential location decisions. Valuation estimates recovered from such a model allow researchers to adjust for an information bias, while allowing for forward-looking behavior in the household's decision problem.

Results suggest that the MWTP for a unit decrease in contamination is \$888.38. This estimate is more than two times higher than the simple hedonic estimate. Although forward-looking behavior have been found in other work to create large biases for valuing other types of neighborhood amenities such as crime and air quality, I find no evidence for households anticipating cleanup from contamination signals once learning about site contamination is allowed. This suggests that as households gain knowledge about site contaminants, they learn that contamination is generally worse than what they had previously thought. Therefore, depending on how household knowledge compares to released information, information that accompanies policies to improve an amenity can significantly alter the researcher's measurement of the MWTP. Furthermore, a breakdown of the MWTP for contamination by future use suggests that households have a much higher willingness to pay for a site with a planned future use. This is something policymakers may want to consider when valuing site cleanup. Since the estimates suggest that households are learning from the information being released, I calculate the value of the information using the parameter estimates from my model. I find the average value per capita for one assessment to be \$17,887.38, where this value varies depending on previous site use. The ranking of the informational value of assessments for different sites is consistent with the notion that there is more value from information released for areas that are least expected to be polluted. In conclusion, these results suggest that the dynamics of information, and in particular, learning, have non-trivial effects on the MWTP estimate, and should be accounted for in recovering amenity valuation.

2.9 Tables

Table 2.1: Brownfield Characteristics

Variable	Obs.	Mean	Median	Std. Dev.	Min.	Max.
Assessments per Site	65	3.43	3	0.98	1	5
NRS_{IV} Score (Institutional)†	65	26.2	15	25.58	0	155
NRS_{V} Score (Environmental)	65	42.2	20	42.82	0	170
Assessment Year	223	2000.08	2002	6.547315	1984	2012
Assessment Interval (yrs)	158	4.51	3	3.759891	0	18
Contaminant (c_{jt})††	223	2.99	3	2.16	0	10
$\Delta c_{jt} = c_{jt} - c_{jt-1}$	158	0.87	1	2.48	-5	7

† NRS Scores IV and V respectively represent measures for the presence institutions and areas of environmental concern.
†† Contaminant (c_{jt}) is the sum of the number of contaminants found in each exposure pathway (soil, groundwater, sediments, air, surface water, or other)

Table 2.2: Site Previous Use

Previous Use Categories †	Freq.	Percent
Construction (G-CON)	3	4.62
Manufacturing (G-MAN)	30	46.15
Natural Resources and Mining (G-NRM)	1	1.54
Educational & Health Services (S-EHS)	3	4.62
Financial Activities (S-FA)	1	1.54
Leisure and Hospitality (S-LH)	2	3.08
Other Services (S-OS)††	11	16.92
Professional & Business Services (S-PBS)	2	3.08
Trade, Transport, Utilities (S-TTU)	10	15.38
Residential (RES)	2	3.08

† Business categories are divided according to the NAICS sector level index.
†† S-OS include Repair & maintenance, Personal & Laundry Services, Religious, Grant-making, Civic, Professional organizations, and Private Households

Table 2.3: Brownfield Contamination

Contaminants

Variable	Obs	Mean	Std. Dev.	Min	Max
Lead	223	0.40	0.49	0	1
Asbestos	223	0.08	0.27	0	1
VOC	223	0.39	0.49	0	1
PCB	223	0.11	0.32	0	1
PAH	223	0.38	0.49	0	1
Petroleum Products	223	0.52	0.50	0	1
Other Metals	223	0.39	0.49	0	1

Exposure Pathways

Variable	Obs	Mean	Std. Dev.	Min	Max
Surface Water	223	0.03	0.16	0	1
Groundwater	223	0.47	0.50	0	1
Soil	223	0.75	0.43	0	1
Other	223	0.12	0.33	0	1

Note: There are a total of 223 signals across 65 brownfield sites.

Table 2.4: Summary Statistics

A. Neighborhood Attributes

Variable	Mean	Median	Stdev.	Min.	Max.
District-level (N = 225 Districts × 14 year)					
Variable	Mean	Median	Stdev.	Min.	Max.
Crime per capita (in 000's)	3.60	2.23	4.09	0.00	36.09
MCAS % Adv/Prof	60.81	65.00	22.54	4.00	100.00
% Low Income	18.43	11.80	18.06	0.00	90.80
Tract-level (N = 1300 Tracts × 14 years)					
Rent (annual)	\$34,838	\$31,866	\$14,506	\$6,315	\$144,754

B. Brownfield Site District Attributes
(N = 65 Sites × 14 years)

Variable	Mean	Median	Stdev.	Min.	Max.
Per Capita Spending (100's)	26.12	25.92	6.26	9.68	48.62
% Unemployed	5.68	5.20	2.48	1.50	17.20
$\mathbf{1}_{[Proposal year]}$	0.07	0.00	0.26	0.00	1.00
% black	0.0522	0.0134	0.1419	0.0003	0.9020
% hispanic	0.0792	0.0256	0.1418	0.0000	0.8337

C. House Attributes

Variable	Mean	Median	Stdev.	Min.	Max.
Transaction Price	\$298,807	\$263,234	\$172,525	\$22,906	\$1,148,206
# of Bedrooms	3.19	3	1.08	1	29
# of Bathrooms	2.10	2	0.82	0.5	32.50
Sq. Feet	1972	1800	923	120	30272
	Total	Movers	Stayers		
Households	678,570	158,319	520,251		

† Since there can be multiple brownfield sites within each tract, NRS scores, which are assigned at the brownfield level, are aggregated up to the tract level.

Table 2.5: Stage 1 Learning Parameter Estimates

		Estimate	95 % Conf. Interv.	
Prior mean	H_0	-1.400		
Prior noise	δ	1.851		
Signal noise	σ_e	0.292		
NRS_{IV}	$\mathbf{IE}_{jt}$	0.033		
NRS_V	$\mathbf{IE}_{jt}$	0.019		
Posterior Beliefs†	Mean	St. Dev.	Min.	Max.
$E_T(H_k)$	-0.480	1.236	-5.067	1.006
$V_T(H_k)$	0.089	0.032	0.057	0.252
	Mean	St. Dev.	Min.	Max.
NRS_{IV} Score (Institutional)	26.2	25.58	0	155
NRS_V Score (Environmental)	42.2	42.82	0	170

† These refer to the most updated set of beliefs by the last time period, T.
†† Contamination has been reparameterized as $\widetilde{c}_{kt} = \log(1 + c_{kt})$

Table 2.6: Stage 2 Discrete Choice Estimates

	Myopic	Forward-looking
Travel Dist (in \$ per km)	\$24.34	\$105.24
Psych MC (in \$)	\$21,589.30	\$10,782.57
Financial MC (in \$1,000)	-0.2367	-0.2419
J × T mean utility estimates	Not Shown	

Note: Dynamic estimate implies an average physical moving cost of \$5,137.90 given an average moving distance of $\overline{distance}$: 48.82 km.

Table 2.7: Tract-Time Mean Utility Decomposition

	Panel A: Myopic Households							
Dep. Var.:	*Assume Full Information*†				*Assume Learning*			
$\widehat{\theta}_{jt}^{myopic} - \widehat{\beta}_{MC}^{myopic} \cdot R_{jt}$	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
sig_{jT^*}	0.001	0.009	-0.086	0.009				
$E[e^{H_j}]$					0.019	0.030	-0.215	0.032
Per capita crime	-90.678	13.396	-46.115	18.460	-91.476	13.444	-41.905	18.561
% Low Income	-0.062	0.002	-0.016	0.008	-0.062	0.002	-0.012	0.008
MCAS % Adv/Prof	0.026	0.002	-0.005	0.007	0.026	0.002	-0.004	0.007
# of Bedrooms	-1.822	0.068	-1.339	0.060	-1.824	0.068	-1.372	0.060
# of Bathrooms	2.953	0.132	2.051	0.121	2.953	0.132	2.054	0.121
Sq. Feet	0.002	0.000	0.001	0.000	0.002	0.000	0.001115	0.000
Fixed Effects	*None*		*District & Year*		*None*		*District & Year*	
	Panel B: Forward-looking Households							
Dep. Var.:	*Assume Full Information*				*Assume Learning*			
$\widehat{\theta}_{jt}^{DDC} - \widehat{\beta}_{MC}^{DDC} \cdot R_{jt}$	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
$sig_{jT^*} \times \widehat{P}_{noclean}$	0.000	0.011	-0.102	0.009				
$E(e^{H_{jt}}) \times \widehat{P}_{noclean}$					-0.099	0.036	-0.217	0.032
Per capita crime	-73.212	15.228	-44.306	17.129	-70.104	15.245	-39.187	17.257
% Low Income	-0.031	0.003	-0.039	0.008	-0.030	0.003	-0.032	0.008
MCAS % Adv/Prof	0.026	0.002	-0.011	0.006	0.025	0.002	-0.010	0.006
# of Bedrooms	-1.996	0.082	-1.254	0.056	-1.986	0.082	-1.300	0.056
# of Bathrooms	3.458	0.156	2.096	0.112	3.450	0.156	2.105	0.112
Sq. Feet	0.002	0.000	0.002	0.000	0.002	0.000	0.002	0.000
Fixed Effects	*None*		*District & Year*		*None*		*District & Year*	

† Under the full information assumption, households are assumed to use the contamination reported from the last assessment for each site before cleanup to make housing choices at all times, as indexed by T^* in the variable *sig*.

Table 2.8: MWTP Estimates

	Panel A: MWTP for Contamination	
Decrease in Contamination	Myopic	Forward-looking
Assume Full Information	\$363.81	\$416.52
	(\$40.09)	(\$134.17)
Assume Learning	\$891.23	\$888.38
	(\$38.85)	(\$130.81)
	Panel B: MWTP for Other Attributes	
Change in Other Attributes	Myopic	Forward-looking
Decrease in Crime per 1000 people	\$177.11	\$165.61
	(\$78.41)	(\$71.35)
Decrease Low Income households by 1% point	\$48.74	\$129.09
	(\$35.27)	(\$34.46)
Decrease in MCAS Adv./Prof. by 1% point	\$15.51	\$41.57
	(\$27.64)	(\$25.36)
Increase in Bedroom	-\$5,796.48	-\$5,313.99
	(\$252.43)	(\$231.31)
Increase in Bathroom	\$8,681.95	\$8,605.98
	(\$510.62)	(\$464.83)
Increase in Sq. Ft.	\$4.71	\$6.71
	(\$0.48)	(\$0.45)

Note: Standard errors in parentheses.

Table 2.9: Simple Hedonic Estimates

Dep. Var. Rental Price	est.	s.e.
sig_{jT^*}	-$373.30	12.901
Crime per 1000 people	-$152.38	15,989.81
% Low Income	$3.69	6.818
MCAS % Adv/Prof	$19.31	3.957
Bed	-$370.10	20.017
Bath	$4,563.00	31.919
Sqft	$10.28	0.03
Constant	$14,689.00	368.696
Observations	147,725	
R-squared	0.634	

Note: Year and district fixed effects included.

Table 2.10: MWTP Estimates by Future Use

Decrease in Contamination	Forward-looking, Learning	
	est.	s.e.
Future Use unknown (Base Group)	$162.01	$92.12
Future Use known	$650.59	$153.90
Future Use: Greenspace	$210.47	$113.63
Future Use: Parking	$107.30	$133.47

Note: Regression includes 3 additional interaction terms with contamination - (1) interaction with dummy for if future use is known, (2) interaction with dummy for whether site will be converted for green space, and (3) interaction with dummy for whether site will be converted for parking.

Table 2.11: Value of Information

A. Overall	Mean	Min	Max
Value Per Household	\$17,887.38	\$8,336.56	\$29,004.13
(Flow Value)	\$894.37	\$416.83	\$1,450.21
# of Tracts within 3km of Sites	20.54	1	76

B. Breakdown by Previous Use†

NAICS Sectors	Mean Value	Mean Flow Value
Professional and Business Services	\$14,089.69	\$704.48
Natural Resources and Mining	\$14,111.16	\$705.56
Residential	\$16,636.31	\$831.82
Trade, Transportation, and Utilities	\$16,654.88	\$832.74
Other Services	\$16,809.43	\$840.47
Educational and Health Services	\$17,768.30	\$888.42
Manufacturing	\$18,778.99	\$938.95
Construction	\$19,884.49	\$994.22
Leisure and Hospitality	\$21,929.50	\$1,096.48
Financial Activities	\$23,307.93	\$1,165.40
C. By Specific Examples Previous Use		
Gas Stations	\$16,240.80	\$812.04
Factories	\$19,209.43	\$960.47
School, Park, Office Building	\$22,771.54	\$1,138.58

Note: The Degroot value of information from removing the last assessment is calculated each of the 65 sites. The table gives summary statistics for the values.
† Previous use categorized by North American Industry Classification System (NAICS)

Table 2.12: Assessment Endogeneity

Dep. Var.	Assessments per Year	Years since last Assessment	Var. Change over Time Mean	St. Dev.
Per capita spending (in $100)	0.00537***	-0.0155	0.910	1.145
	(0.001)	(0.028)		
% Unemployed	0.00579**	0.0605	0.357	1.104
	(0.002)	(0.053)		
% Low Income	0.000193	-0.0220**	0.900	2.680
	(0.000)	(0.010)		
MCAS % Adv/Prof	0.00117***	-0.0172**	3.901	7.739
	(0.000)	(0.008)		
Per capita crime	-2.559	32.68	0.000	0.001
	(1.767)	(39.317)		
Constant	-0.176***	4.379***		
	(0.025)	(0.555)		
Obs.	786	786		
Dep. Var Mean	0.0547	2.8270		

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. Dependent variables are calculated for each brownfield site at each year between 1998 and 2011. All independent variables vary at the school-district and year level.

Table 2.13: Average Price Change at Various Assessment Updates

Dep. Var. log(Price)	First Assmt	Assmt 2	Assmt 3	Assmt 4	Assmt 5	Last Assmt
$1_{[\text{After First Assmt}]}$	-0.166***					
	(0.016)					
$1_{[\text{After Assmt 2}]}$		-0.00481				
		(0.008)				
$1_{[\text{After Assmt 3}]}$			-0.0577***			
			(0.007)			
$1_{[\text{After Assmt 4}]}$				-0.0175*		
				(0.010)		
$1_{[\text{After Assmt 5}]}$					-0.145***	
					(0.027)	
$1_{[\text{After Last Assmt}]}$						0.00687
						(0.005)
Constant	10.98***	11.36***	11.22***	11.29***	12.03***	11.26***
	(0.080)	(0.048)	(0.048)	(0.171)	(0.163)	(0.077)
Observations	14,651	25,973	35,981	23,260	9,381	57,718
R-squared	0.528	0.535	0.540	0.511	0.566	0.550

Note: Each column represents a regression of log(price) on a dummy variable that takes a value of 1 if a house sells after the corresponding assessment. The base group for comparison are houses that sold before the assessment. Observations are limited to 730 days (approx. 2 years) before and after each assessment. Regressions control for house attributes (square footage, number of bedrooms and bathrooms), IE settings, contamination level before assessments released, as well as year and district fixed effects. Transactions are limited to houses within 3km of a brownfield site. Standard errors in parentheses. Standard errors are in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 2.14: CCP Approximation

	Utility Parameters		
Travel Dist	-0.0082		
Psych MC	-3.8215		
Financial MC	-0.2853		
	Mean Utility Forecast		
θ_{jt-1}	0.5464		
Time Period	-0.9064		
Dummy Variables	-1.2785		
	-0.1201		
	-1.2873		
	-2.0402		
	-1.5651		
	$E[logPr(d_{it+1} = 1 \| d_{it})]$		
	Mean	Min	Max
N-by-J matrix	-9.0399	-31.0834	-0.8658
# of Draws	25		

Note: 25 draws from distribution of residuals from the mean utility forecast were taken to form the expected log probability of choosing k conditional on choice at t.

Table 2.15: Binary Logit for Cleanup

	est.	s.e.	p-value
# of assessments	0.4310	0.1351	0.0014
GW contamination	-0.5002	0.3575	0.1617
Spending per Capita (in $100)	0.1307	0.0309	0.0000
Unemployment Rate	0.0624	0.0701	0.3736
1[Proposal Year]	1.6840	0.4035	0.0000
% Black (non-time varying)	-1.7548	1.0994	0.1105
% Hispanic (non-time varying)	-3.3975	1.3301	0.0106
Signal	0.0062	0.0814	0.9396
Constant	-7.03	0.88	0.00
N	683		

Note: Cleanup probabilities are predicted at each time period, for each site. Neighborhood-level characteristics that may affect cleanup probability are taken for the tract or district that contains the site.

2.10 Figures

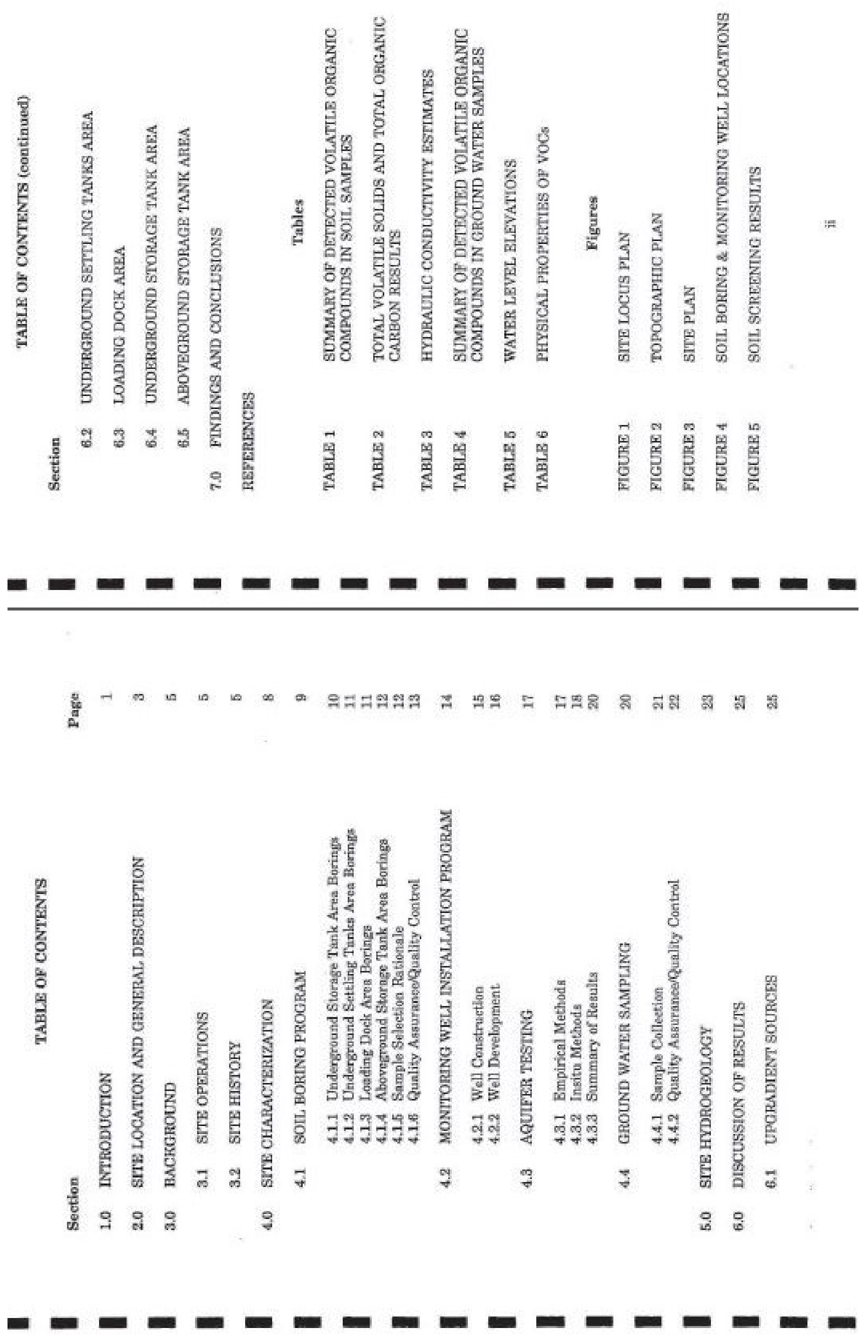


TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

ii

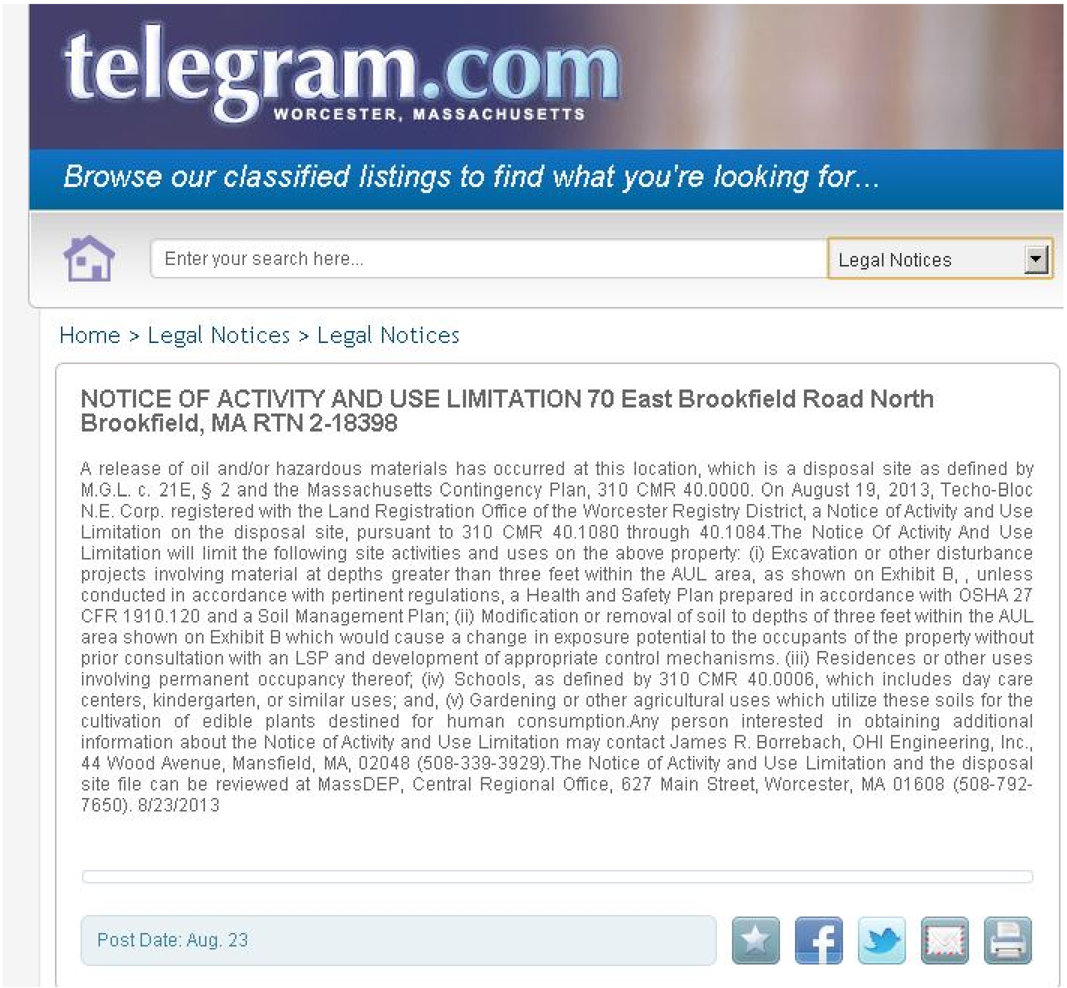
FIGURE 2.1: Assessment Example (Contents)

SUMMARY OF PHASE II SOIL ANALYTICAL RESULTS
PHASE II COMPREHENSIVE SITE ASSESSMENT REPORT
FORMER NEW ENGLAND LOG HOMES SITE
100 BRIDGE STREET, GREAT BARRINGTON, MASSACHUSETTS
JANUARY 2002

PARAMETERS	Sample Location	SB-21/MW-11	SB-23/MW-12	SB-10A/MW-13	SB-22	SB-24	SB-24	SB-25	SB-25	SB-25	SB-26	SB-27			SB-28	
	F&O Sample #	111001128-08	111001128-12	111001128-21	111001128-09	111001128-22	111001128-23	111001128-01	111001128-02	111001128-03	111001128-04	111001128-05	111001128-06	111001128-07	111001128-16	111001128-17
	Depth (feet)	1-3	1-3	18-20	1-3	1-3	10-12	1-3	1-3 (Duplicate)	10-12	0.25-1.25	1-3	5-7	10-12	1-3	10-12
	Date	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00	11-28-00
Extractable Petroleum Hydrocarbons (EPH)	(ug/kg)															
C9-C18 Aliphatics		---	---	10000	---	---	5900	---	---	---	<3200	---	---	---	<3800	4700
C19-C36 Aliphatics		---	---	<4600	---	---	5600	---	---	---	110000	---	---	---	10000	<4900
C10-C22Aromatics		---	---	<9700	---	---	<11000	---	---	---	30000	---	---	---	<11000	<10000
EPH Target Analytes (PAHs)	(ug/kg)															
Naphthalene		<380 U	<390 U	---	<380 U	(370) J	---	<360 U	---	---	<360 U	(90) J	---	---	<410 U	<420 U
2-Methylnaphthalene		<380 U	<390 U	---	<380 U	(440) J	---	<360 U	---	---	<360 U	(59) J	---	---	<410 U	<420 U
Acenaphthylene		<380 U	<390 U	---	(45) J	2300	---	<360 U	---	---	<360 U	(230) J	---	---	<410 U	<420 U
Acenaphthene		(47) J	<390 U	---	<380 U	(200) J	---	<360 U	---	---	<360 U	(220) J	---	---	<410 U	<420 U
Dibenzofuran		<380 U	<390 U	---	<380 U	(190) J	---	<360 U	---	---	<360 U	(250) J	---	---	<410 U	<420 U
Fluorene		<380 U	<390 U	---	<380 U	(530) J	---	<360 U	---	---	<360 U	(280) J	---	---	<410 U	<420 U
Phenanthrene		580	(200) J	---	440	6100	---	<360 U	---	---	<360 U	4300	---	---	<410 U	<420 U
Anthracene		(110) J	(43) J	---	(94) J	2400	---	<360 U	---	---	<360 U	980	---	---	<410 U	<420 U
Fluoranthene		830	(190) J	---	690	10000	---	<360 U	---	---	<360 U	5500	---	---	<410 U	<420 U
Pyrene		760	(180) J	---	650	11000	---	<360 U	---	---	<360 U	5100	---	---	<410 U	<420 U
Benzo(a)anthracene		380	(95) J	---	(350) J	5500	---	<360 U	---	---	<360 U	2400	---	---	<410 U	<420 U
Chrysene		410	(84) J	---	430	6300	---	<360 U	---	---	<360 U	2500	---	---	<410 U	<420 U
Benzo(b)fluoranthene		410	(100) J	---	440	6600	---	<360 U	---	---	<360 U	2700	---	---	<410 U	<420 U
Benzo(k)fluoranthene		(140) J	(110) J	---	(130) J	2100	---	<360 U	---	---	<360 U	960	---	---	<410 U	<420 U
Benzo(a)pyrene		(350) J	(74) J	---	(350) J	4900	---	<360 U	---	---	<360 U	2200	---	---	<410 U	<420 U
Indeno(1,2,3-cd)pyrene		(180) J	<390 U	---	(190) J	3000	---	<360 U	---	---	<360 U	1300	---	---	<410 U	<420 U
Dibenzo(a,h)anthracene		<380 U	<390 U	---	<380 U	(850) J	---	<360 U	---	---	<360 U	(360) J	---	---	<410 U	<420 U
Benzo(ghi)perylene		(200) J	<390 U	---	(230) J	2700	---	<360 U	---	---	<360 U	1200	---	---	<410 U	<420 U
Pentachlorophenol	(ug/kg)	---	---	2500	---	---	<850 U	---	---	<890 U	---	(120) J	<800 U	(76) J	---	---
Metals	(mg/kg)															
Arsenic		4.5	---	---	11.7	39.7	---	12.7	7.6	---	---	6.6	---	---	---	---
Beryllium		---	---	---	---	---	---	0.22 B	<0.14 U	---	---	0.49 B	---	---	---	---
Lead		21.1	10	---	58	---	---	17.4	10.1	6.9	---	---	---	---	---	---
Zinc		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Dioxin (Method 8290)	(ng/kg)															
2,3,7,8-TCDD TE		---	---	---	---	---	---	---	---	---	310	---	---	---	---	---
2,3,7,8-Tetrachlorodibenzofuran		---	---	---	---	---	---	---	---	---	1.1	---	---	---	---	---
2,3,7,8-TCDD (Dioxin)		---	---	---	---	---	---	---	---	---	<0.50	---	---	---	---	---
1,2,3,7,8-Pentachlorodibenzofuran		---	---	---	---	---	---	---	---	---	6.1	---	---	---	---	---
2,3,4,7,8-Pentachlorodibenzofuran		---	---	---	---	---	---	---	---	---	8.8	---	---	---	---	---
1,2,3,7,8-Pentachlorodibenzo-p-dioxin		---	---	---	---	---	---	---	---	---	4.9 J	---	---	---	---	---
1,2,3,4,7,8-Hexachlorodibenzofuran		---	---	---	---	---	---	---	---	---	14	---	---	---	---	---
1,2,3,6,7,8-Hexachlorodibenzofuran		---	---	---	---	---	---	---	---	---	74 Q	---	---	---	---	---
2,3,4,6,7,8-Hexachlorodibenzofuran		---	---	---	---	---	---	---	---	---	26	---	---	---	---	---
1,2,3,7,8,9-HxCDF (For TOTAL HCDF)		---	---	---	---	---	---	---	---	---	8.9	---	---	---	---	---
1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin		---	---	---	---	---	---	---	---	---	13	---	---	---	---	---
1,2,3,6,7,8-Hexachlorodibenzo-p-dioxin		---	---	---	---	---	---	---	---	---	94	---	---	---	---	---
1,2,3,7,8,9-Hexachlorodibenzo-p-dioxin		---	---	---	---	---	---	---	---	---	31	---	---	---	---	---
1,2,3,4,6,7,8-Heptachlorodibenzofuran		---	---	---	---	---	---	---	---	---	290	---	---	---	---	---
1,2,3,4,7,8,9-Heptachlorodibenzofuran		---	---	---	---	---	---	---	---	---	22	---	---	---	---	---
1,2,3,4,6,7,8-HpCDD		---	---	---	---	---	---	---	---	---	2000	---	---	---	---	---
Octachlorodibenzofuran		---	---	---	---	---	---	---	---	---	770	---	---	---	---	---
Octachlorodibenzo-p-dioxin		---	---	---	---	---	---	---	---	---	17000	---	---	---	---	---
TOTAL TCDF		---	---	---	---	---	---	---	---	---	11	---	---	---	---	---
TOTAL TCDD		---	---	---	---	---	---	---	---	---	<0.50	---	---	---	---	---
TOTAL PeCDF		---	---	---	---	---	---	---	---	---	83	---	---	---	---	---
TOTAL PeCDD		---	---	---	---	---	---	---	---	---	13	---	---	---	---	---
TOTAL HxCDF		---	---	---	---	---	---	---	---	---	540	---	---	---	---	---
HxCDDs (total)		---	---	---	---	---	---	---	---	---	350	---	---	---	---	---
HpCDFs (total)		---	---	---	---	---	---	---	---	---	960	---	---	---	---	---
HpCDDs (total)		---	---	---	---	---	---	---	---	---	3200	---	---	---	---	---

NOTES:
B = Constituent found in blank as well as sample
J = Estimated value
( ) = Value less than Method Detection Limit, MDL
( ) J = Estimated value less than MDL
Q = Estimated maximum possible concentration (generally due to interference problems)
U = Not detected above reported MDL

FIGURE 2.2: Assessment Example (Soil Boring Results)


telegram.com
WORCESTER, MASSACHUSETTS

Browse our classified listings to find what you're looking for...

Enter your search here...

Legal Notices

Home > Legal Notices > Legal Notices

NOTICE OF ACTIVITY AND USE LIMITATION 70 East Brookfield Road North Brookfield, MA RTN 2-18398

A release of oil and/or hazardous materials has occurred at this location, which is a disposal site as defined by M.G.L. c. 21E, § 2 and the Massachusetts Contingency Plan, 310 CMR 40.0000. On August 19, 2013, Techo-Bloc N.E. Corp. registered with the Land Registration Office of the Worcester Registry District, a Notice of Activity and Use Limitation on the disposal site, pursuant to 310 CMR 40.1080 through 40.1084.The Notice Of Activity And Use Limitation will limit the following site activities and uses on the above property: (i) Excavation or other disturbance projects involving material at depths greater than three feet within the AUL area, as shown on Exhibit B, , unless conducted in accordance with pertinent regulations, a Health and Safety Plan prepared in accordance with OSHA 27 CFR 1910.120 and a Soil Management Plan; (ii) Modification or removal of soil to depths of three feet within the AUL area shown on Exhibit B which would cause a change in exposure potential to the occupants of the property without prior consultation with an LSP and development of appropriate control mechanisms. (iii) Residences or other uses involving permanent occupancy thereof; (iv) Schools, as defined by 310 CMR 40.0006, which includes day care centers, kindergarten, or similar uses; and, (v) Gardening or other agricultural uses which utilize these soils for the cultivation of edible plants destined for human consumption.Any person interested in obtaining additional information about the Notice of Activity and Use Limitation may contact James R. Borrebach, OHI Engineering, Inc., 44 Wood Avenue, Mansfield, MA, 02048 (508-339-3929).The Notice of Activity and Use Limitation and the disposal site file can be reviewed at MassDEP, Central Regional Office, 627 Main Street, Worcester, MA 01608 (508-792-7650). 8/23/2013

Post Date: Aug. 23

Figure 2.3: Public Notification

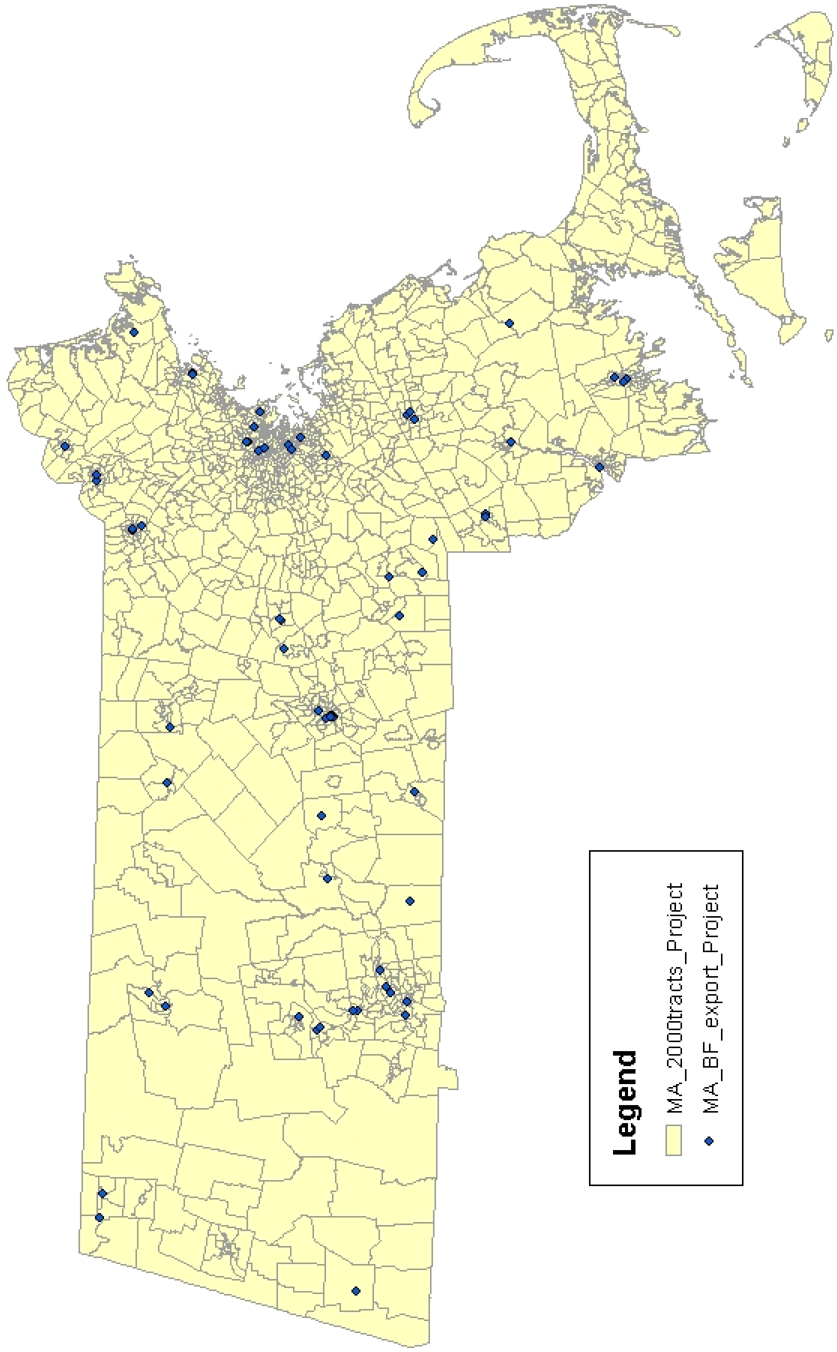


Figure 2.4: Brownfields in Massachusetts

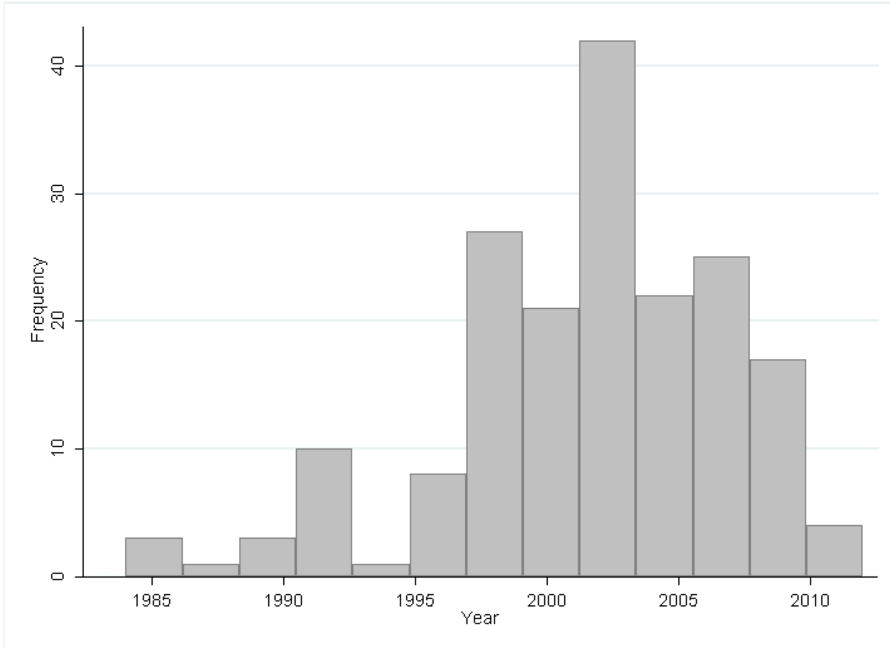


Figure 2.5: Distribution of Assessments over Time

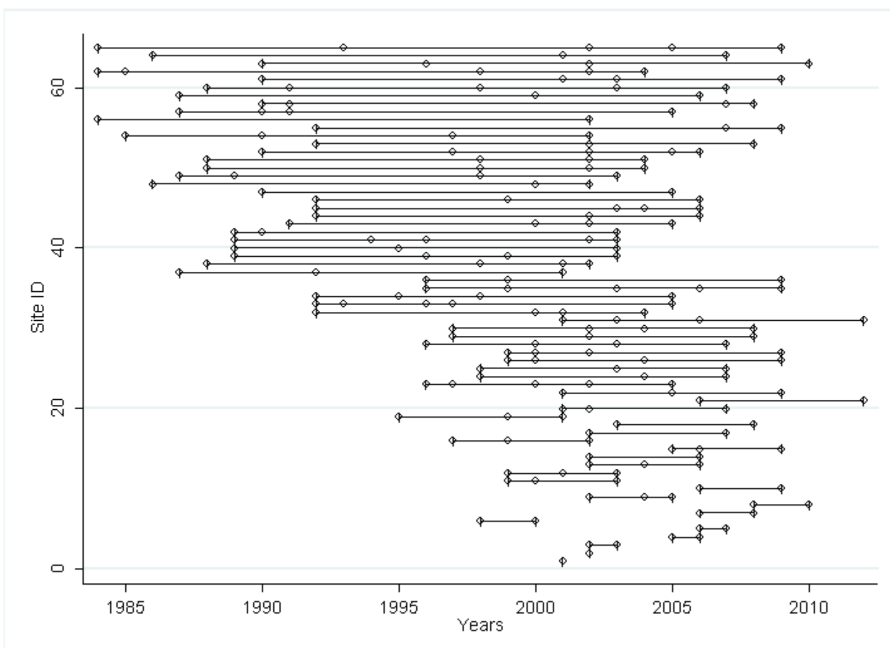


Figure 2.6: Assessment Interval (in Years)

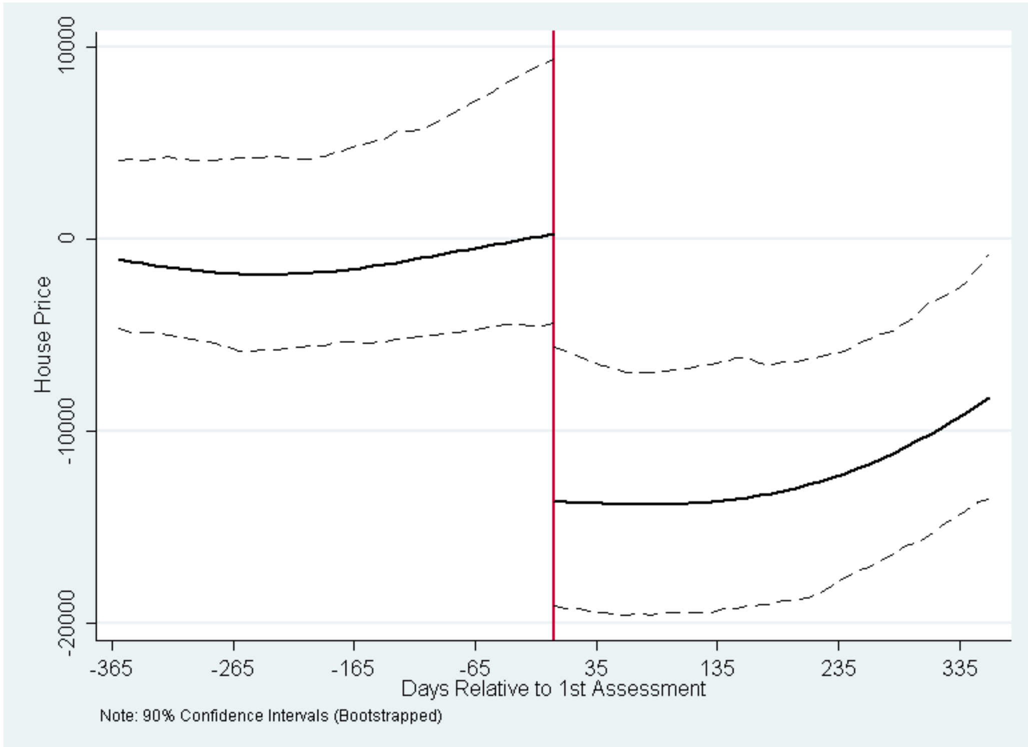


Figure 2.7: Housing Prices 2 Years Before and After 1st Assessment

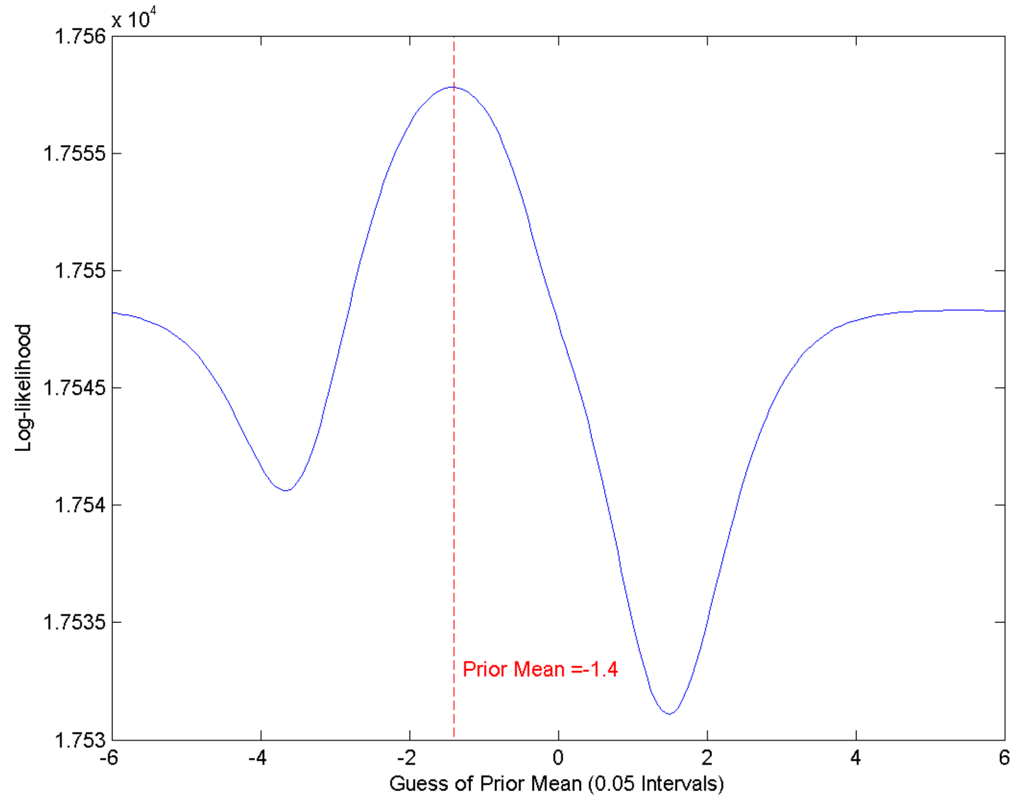


FIGURE 2.8: Log-likelihood Plot for Prior Mean Grid Search

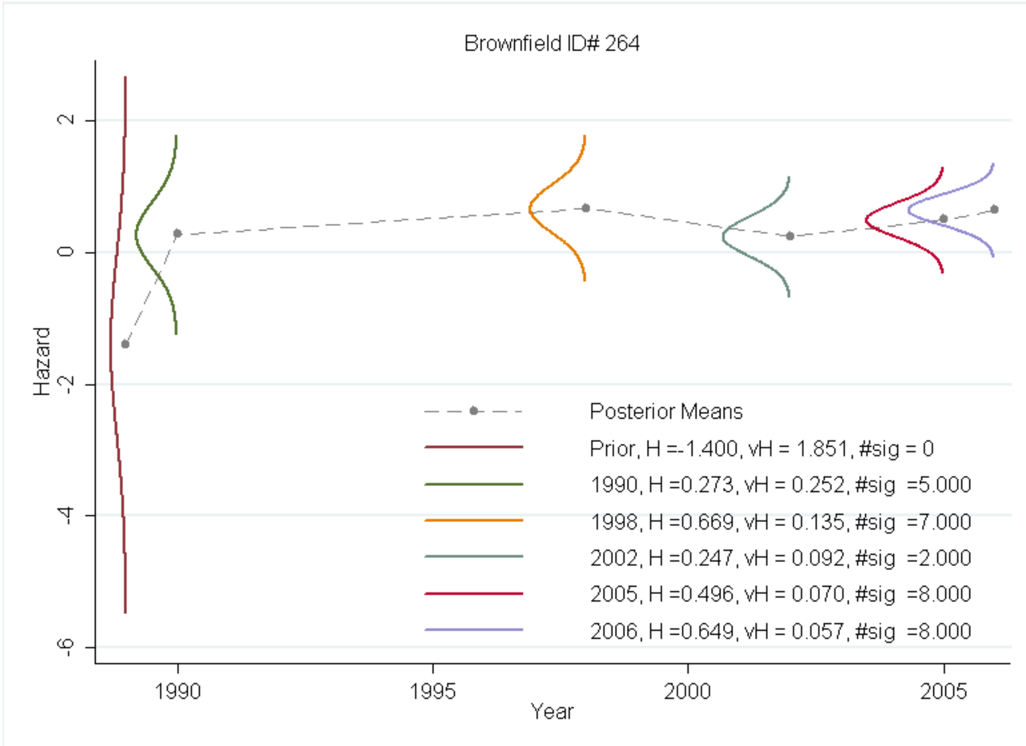


FIGURE 2.9: Posterior Belief at each Signal Update

2.11 Appendix: EM Alogirthm

2.11.1 *Expectation-step*

The E-step involves estimating the most updated posterior beliefs on the hazard (summarized by the distribution $F(H)$) given a guess at the learning parameters, and then building the likelihood of observing the contamination results using the estimated hazard belief (distribution). Since I assume that learning for each brownfield (of each neighborhood) occurs independently, the most updated posteriors at the p^{th} iteration, $E_T(H)^{(p)}$ and $V_T(H)^{(p)}$, can be found for each brownfield k using the current guess at the parameters, $\theta^{(p)}$, with the following Bayesian updating formulas

$$\begin{aligned} E_T(H_k)^{(p)} &= \left[\frac{1}{\delta^{(p)}} + \frac{N_k}{\sigma_e^{(p)}}\right]^{-1} \cdot \left(\frac{1}{\sigma_e^{(p)}} \sum_{t=1}^{N_k} \left(c_{kt} - \lambda^{(p)} \cdot IE_k\right) + \frac{1}{\delta^{(p)}} H_0^{guess}\right) \\ V_T(H_k)^{(p)} &= \left[\frac{1}{\delta^{(p)}} + \frac{N_k}{\sigma_e^{(p)}}\right]^{-1} \end{aligned}$$

The above formulas essentially calculate a weighted average of the signals and the given prior for each brownfield, where N_k denotes the total number of assessments performed for a brownfield k. This recovers a hazard level for each brownfield, which is also our best guess of each site's hazard given the string of information updates.[24] Given the most updated posteriors on the hazard, we can build the log-likelihood of observing the contamination by integrating over the log probability of the observed contamination with respect to the hazard.

2.11.2 *Maximization-step*

In the M-step, the likelihood from the E-step is maximized with respect to the learning parameters to produce updated learning parameters, $\left[\lambda^{(p+1)}, \sigma_e^{(p+1)}, \delta^{(p+1)}\right]$.

[24]There is a concern that for sites that undergo very few assessments (e.g. 1), our 'best' guess is not very good.

Using the updated parameters, $\left[\lambda^{(p+1)}, \sigma_e^{(p+1)}, \delta^{(p+1)}\right]$, I return to the E-step and re-estimate the posterior distribution on the hazard level. Dempster et al. (1977) show that by iterating between the two steps, the estimates will eventually converge.

Using the Bayesian updating formulas, the final learning parameter estimates given the guess of the prior, and the string of signals up until time t, we can construct the beliefs about H_k at time t for each t in our time frame. I combine beliefs for multiple brownfields within a neighborhood, j, by summing the belief distributions. The resulting posterior beliefs for brownfield hazards in a neighborhood j, with K_j brownfields, would be given by $N\left(E_t[H_1] + \cdots + E_t[H_{K_j}], V_t[H_1] + \cdots + V_t[H_{K_j}]\right)$.[25]

[25]This is since the sum of independent normal random variables is normal, and the learning is assumed to be independent between brownfields, e.g. learning about one site does not provide information about another.

3

Exceeding the Threshold: Hedonic Analysis of Brownfield Contamination

with Gabrielle Inder[1]

3.1 Introduction

Public information on environmental quality is necessary for market-based polices aimed at correcting environmental externalities to work. As these policies only function to the extent that households are aware of the pollution they face, provision of environmental information is important from a policy perspective for two reasons. First, improved information on pollution may allow households to adjust their behavior and make more informed decisions on mitigating pollution exposure (Viscusi et al. (1986), Smith and Johnson (1988), Graff Zivin et al. (2011), Graff Zivin and Neidell (2009)). Second, information provision can be a source of environmental regulation by inducing polluting firms to reduce their pollution in response to household demand for improved environmental quality (Hamilton (1995), Konar and Cohen (1997), Khanna and Damon (1999), Powers (2013)). One of the main objectives of the Emergency Planning and Community Right-to-Know Act (EPCRA) (EPA

[1]Inder: Duke University Department of Economics

(2014)), enacted in 1986, was to established reporting requirements for government agencies and industry that handled hazardous and toxic chemicals so as to increase public knowledge and access to pollution information, and eventually reduce pollution.

Because of this interest in information provision, there has been a wave of research examining whether information or publicity about contamination is capitalized into housing prices as a way of either testing whether households demand for environmental quality or whether households adjust their risk perceptions in the face of changing information (McClelland et al. (1990), Gayer and Viscusi (2002), Bui and Mayer (2003), Oberholzer-Gee and Mitsunari (2006), Sanders (2013), Decker et al. (2005)). Most of these papers find that information is capitalized into property values negatively (postively) if households adjust risk perceptions upward (downward) after information release. It is not surpising that, although findings are generally consistent, estimates vary as multiple ways of quantifying 'information' have been used. These measures include (but are not limited to) public administrative data on toxins released (Bui and Mayer (2003), Oberholzer-Gee and Mitsunari (2006), Sanders (2013), Decker et al. (2005)), newspaper coverage (Gayer and Viscusi (2002)), direct solicitation of household perceived information (Gawande and Jenkins-Smith (2001)), and environmental site or risk assessments (McClelland et al. (1990), Ma (2014)). One reason for the range in estimates may simply be that people pay attention to different forms of information differently. Bui and Mayer (2003) address this possibility in their study of how information release from the Toxic Release Inventory (TRI) program impacts housing prices. In particular, they do not find that increasing information provision affects housing prices, and attribute their findings to the lack of ability for individuals to process the pollution information provided. If the form of the information released through the TRI program is too complex, then would characterizing information in other ways improve the chances of that information reaching households? From a policy perspective, shedding light on how the type of information release can vary in its impact on household choices can be informative as it may allow for better design of information provision programs aimed to inform the public and improve pollution.

This paper aims to contribute to the literature in environmental information provision by examining whether different types of information about brownfield contamination capitalize into property values differently. More specifically, we examine how housing values are impacted if information about contamination is released as a continuous measure as opposed to a binary measure (i.e. exceeding a contamination threshold value or not). We do this by exploiting variation in contaminant thresholds used, holding constant the contaminant level, due to regulatory requirements for brownfield investigations in the state of Massachusetts. As the variation in threshold levels are tied to the level of human exposure of the areas in which these contaminated sites exist, threshold exceedance is potentially correlated with unobserved neighborhood characteristics that also impact housing values. To deal with this, we take an instrumental variables approach using variation in threshold exceedance due to the location of underground water sources. Section 3.2 introduces our study context using brownfields, and the reporting framework for contamination that we exploit. Section 3.3 outlines a hedonic model, and discusses potential endogeneity issues related to our variables of interest. Section 3.4 examines the validity of our instrumental variable, and gives the results. Finally, section 3.5 concludes.

3.2 Study Context

This paper examines information provision for brownfields in the state of Massachusetts. The set of brownfields considered include a subset of the brownfields tied to the US EPA Brownfields Program. In the following section, we begin with a description of brownfields and the federal Brownfields Program. We then review the reporting requirements on contamination that are specific to the state of Massachusetts, which allow us to test whether different information provision schemes capitalize into housing values differently.

3.2.1 The Brownfields Program

While Massachusetts has no formal definition of a brownfield site, the U.S. Environmental Protection Agency (EPA) defines a brownfield as a 'real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant' (EPA (2012)). In most cases, brownfield lands are the result of businesses engaging in industrial or commercial activity that involved the use of toxic substances. The U.S. Government Accountability Office (GAO) estimates that there are over 450,000 brownfield lands in the U.S., and potentially more that have not been 'discovered' or recognized. To encourage cleanup and redevelopment of these sites, the Small Business Liability Relief and Brownfields Revitalization Act was signed in 2002. The law relieved liability concerns for potential future owners by amending the provisions of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980, and allocated grant funding for programs related to brownfield remediation. There are four types of grants designated: assessment grants, job training grants, cleanup grants and revolving loan fund grants. Each grant serves a different purpose for the community; however all represent important aspects of remediation efforts. These grants typically give up to $200,000 for their specified purpose. Assessment grants are the most common type of grant awarded, providing funds to be used for planning of a brownfield cleanup. On the other hand, job training grant funds are typically the least common, and are used to find and train unemployed and low-income residents from brownfield- affected areas to assist with the site cleanup. Cleanup grants remediate brownfields that have been exposed to petroleum or hazardous pollutant contamination. Finally, revolving loan fund grants help to capitalize a revolving loan fund, providing loans and sub-grants for clean up processes. Between 2002 and 2011, the U.S EPA awarded over $670 million for Brownfield grants.[2] The set of brownfields considered in this study include those for which a cleanup grant application was submitted to the EPA from Massachusetts.

3.2.2 Contamination Reporting in Massachusetts

Before any cleanup efforts can commence, a brownfield site must be assessed for its contaminants. The Massachusetts Department of Environmental Protection (MA-DEP) follows a set of guidelines outlined in the Massachusetts Contingency Plan (MCP), which details a formal process for investigating and remediating contaminated lands. The MCP requires site assessments be performed by licensed professionals at the discovery of potential contamination. For each assessment, samples of soil, groundwater, air, and sediments are taken, and tested for suspected hazardous substances. The amounts of the toxins (e.g. microgram per liter (ug/l) for surface/groundwater or microgram per gram (ug/g) for soil/sediments) are then compared to a threshold level for that specific toxin that is considered hazardous for humans. The thresholds with which to compare toxin levels are lower if there are higher risks of human exposure (e.g. school) or environmental damages (e.g. wetland). Specifically in Massachusetts, these standards are defined as GW-1 through GW-3 for groundwater, with GW-1 being the highest threshold, and S-1 through S-3 for soil, with S-1 being the highest threshold. The MA-DEP uses specific criteria for human population and land uses near each property to categorize sites and determine which threshold standards to use for evaluating pollution risk. For human populations, it examines whether there is residential population within a mile, institutions (e.g. schools, hospitals, community centers) within 500 feet, and on-site workers. For

[2] $331.3 million for 1,479 assessment grants, $25.2 million for 121 job training grants, $150.7 million for 801 cleanup grants, and $167.5 million for 143 revolving loan grants

land use, more weight is given to sites that are near a sole source aquifer, potentially productive aquifer, or within 500 feet of a drinking water source, especially where no alternate water supply exists. Given this method of evaluating contamination, it is possible to observe two brownfield sites with the same amount of toxicity, but one exceeds a contamination threshold and the other does not due to the environment around the site. This creates a situation in which, with the appropriate econometric methods, we can test to see if households care about exceeding the threshold for contamination while holding toxicity constant, or the actual amounts of the toxicity. In particular, are threshold violations capitalized by the housing market.

3.3 Model

Rosen (1974) provided the theoretical foundations for using property transaction prices to reveal household willingness to pay for non-marketed amenities. We follow the previous literature by examining threshold effects within a property value hedonic model. To examine whether exceeding a threshold for contamination affects housing prices, while holding the level of toxicity constant, we model the price of a property that is located in school district, j, near brownfield site, m, as the following

$$P_{ijmt} = \alpha_0 + \alpha_1 \cdot maxtox_{mt} + \alpha_1 \cdot exceed_{mt} + \epsilon_{ijmt} \tag{3.1}$$

where the dependent variable, P_{ijmt}, is the log price for house i in district j near brownfield m at time t, and ϵ_{ijmt} is an idiosyncratic shock that is specific to the property at time t. The variables of interest are $maxtox_{mt}$, which refers to the toxicity found at brownfield m at time t, and $exceed_{mt}$, which is a dummy variable that equals to 1 if the amount of toxicity found at t exceeds a threshold value that is deemed hazardous. We expand the model to control for other factors that may influence housing prices, including property structural characteristics, X_{it}, observable time-varying district-level characteristics, NBD_{jt}, and characteristics of the nearest brownfield site that do not vary with time, BF_m

$$P_{ijmt} = \alpha_0 + \alpha_1 \cdot maxtox_{mt} + \alpha_1 \cdot exceed_{mt} + X_{it}\beta + NBD_{jt}\gamma + BF_m\delta + \epsilon_{ijmt} \quad (3.2)$$

Structural characteristics include the number of bathrooms and bedrooms, square footage, age, and dummy variables for whether the property is a condo or a single family house. Time-varying neighborhood attributes include measures of poverty, crime, and education. Time-invariant brownfield site characteristics include whether a site is eventually awarded with cleanup, the proposal score, the grant application type (petroleum, hazardous substances, or both), property size, whether there has ever been industrial activity, residential dwellings, or greenspace nearby, and whether there are schools within 500 meters of the brownfield property.

Beyond these observable attributes for which we have data, there may still be unobserved aspects of residential properties that are correlated with our variables of interest and affect housing prices. First, if certain brownfield neighborhoods are located in historically more industrial communities that are also likely to have higher contamination levels, then we might overstate the negative impact of contamination on housing prices. Second, as described in the previous section, contamination exceedance depends on the local environment around the site, and thus may be correlated with attributes of the neighborhood around the brownfield that are unobserved by the researcher. In particular, brownfield sites with more chances of human exposure have higher chances of exceeding thresholds (holding the level of contamination constant), but these sites may also be the ones that are located near more amenities. If local amenities raise housing prices, then one would expect the presence of these unobserved amenities to diminish threshold exceedance's negative effect on housing prices. We decompose the idiosyncratic error term to include district-specific factors, μ_j, and brownfield-specific factors, ν_{mt},

$$\epsilon_{ijmt} = \mu_j + \nu_{mt} + \eta_{ijmt} \quad (3.3)$$

To deal with unobserved time-invariant district attributes, we use school district-level fixed effects and demean the model in (3.2)

$$\begin{aligned} \overline{P}_{ijmt} = \alpha_1 \cdot \overline{maxtox}_{mt} \quad &+ \quad \alpha_1 \cdot \overline{exceed}_{mt} \\ &+ \quad \overline{X}_{it}\beta + \overline{NBD}_{jt}\gamma + \overline{BF}_{m}\delta + \overline{\nu}_{mt} + \overline{\eta}_{ijmt} \end{aligned} \tag{3.4}$$

where $\overline{Y}_{ijmt} = Y_{ijmt} - \sum_{i,m,t \in j} Y_{ijmt}$. Using the within-estimator removes the time-invariant factors at the district level, μ_j, that are correlated with the amount of contamination found at sites, $maxtox_{mt}$.

To deal with endogeneity related to threshold exceedance, we employ an instrumental variables strategy based on site assessment rules for thresholds in Massachusetts. As briefly mentioned in the preceding section, the presence of underground water sources are used to select the appropriate thresholds with which to compare contaminant levels in addition to proximity to human institutions. If the presence of underground water sources are uncorrelated with the institutions that also serve as neighborhood amenities, then we could induce variation in the likelihood of exceeding a threshold while holding constant the unobserved factors $\overline{\nu}_{mt}$, and consistently estimate α_1. Specifically, we use whether a brownfield site is located on top of an aquifer, that is,

$$aquifer_m = \begin{cases} 1 & \text{if brownfield to aquifer distance} = 0 \\ 0 & \text{otherwise} \end{cases}$$

The geographic variation provided by aquifer location may better control for institutional factors than school district fixed effects because it is at a finer level of geography, at which threshold exceedances for brownfield sites are more likely determined. For $aquifer_m$ to be a valid instrument, it must be the case that (1) being situated above an aquifer leads to a higher chance of brownfield contamination levels exceeding thresholds (instrument relevance), or

$$E\left(\overline{aquifer}_m \cdot \overline{exceed}_{mt} \mid maxtox_{mt}, X_{it}, NBD_{jt}, BF_m\right) \neq 0 \tag{3.5}$$

and (2) aquifer location is not correlated with other unobserved factors that affect housing prices (instrument exogeneity) so that after controlling for district-level unobservables,

$$E\big(\overline{aquifer}_m \cdot \overline{\nu}_{mt} \mid maxtox_{mt}, X_{it}, NBD_{jt}, BF_m\big) = 0 \tag{3.6}$$

As the exogeneity condition is not formally testable, one can look to see if significant differences still exist for observable house attributes after conditioning on the instrument. Of special concern for the exogeneity of the instrument is whether the presence of an aquifer is correlated with whether residences are groundwater or piped-water dependent for a source of drinking water. As groundwater dependent houses may be located in more suburban areas, locating on an aquifer may negatively impact housing prices outside of its effect through threshold exceedance. Unfortunately, because public data on geographic coverage for Piped Water Service Areas (PWSAs) in Massachusetts is incomplete, we cannot currently control for this omitted variable. Going forward, we plan on working to obtain this data on PWSAs. We discuss the results from 'balance tests' (Chay and Greenstone (2005)) to assess the validity of the instrument in Section 5.

3.4 Data and Summary Statistics

3.4.1 Brownfields Toxicity and Threshold Data

The set of brownfield sites considered are those that submitted applications to the EPA for a cleanup grant from 2003 to 2008. Figure 3.1 maps the brownfield sites in Massachusetts. Data on contamination from brownfield site investigations are collected from assessment reports that were compiled by Licensed Environmental Site Professionals.[3] For this analysis, information on the amount of contaminants and threshold levels were collected.

[3]For more detailed information about these reports, see Ma (2014)

Most assessments collected pollutant concentrations for soil and groundwater, while air and sediment data was sporadically collected when a perceived threat was in existence. The majority of sites collected soil and groundwater from 4 to 10 locations around the site to determine differences in pollutant concentrations. Multiple samples of the same pollutant can be collected, and the maximum detected concentration for each pollutant was recorded for this data set. Since different chemicals have different safety level concentrations, concentration values were assessed against their corresponding threshold values to make the toxicity comparable across chemicals. Specifically, concentration is divided by the threshold value and multiplied by 100, giving the relative danger of that contaminant. This approach implies that any toxicity over 100 indicates an exceedance of the threshold for that given contaminant.

For the 66 brownfield sites analyzed, 17 have yet to be cleaned, while the majority have undergone some form of remediation. In total, 293 unique chemicals were listed as found across all sites, which can be classified into 20 groups, including post transition metals, transition metals, xylenes, ringed polyaromatic hydrocarbons and napthalenes.[4] See Table 3.1 for a list of chemical groups, and Table 3.2 for a list of chemicals found. Approximately 9,000 data points were recorded as the maximum concentrations of unique chemicals at all sites in all mediums. As many of the chemicals are complements in industrial/commercial production processes, rather than trying to distinguish between the effects of exceeding each of the 293 chemical thresholds, the maximum toxicity value was taken for each investigation, regardless of the chemical, to avoid issues relating to multicollinearity. An investigation finds a threshold exceedance if at least one chemical exceeds its threshold value. Table 3.3 gives summary statistics for site investigations.

[4]It is important to note that every site was not tested for every chemical. This is because some chemicals, such as pesticides, are only produced or used in certain industrial sites, and thus would not be present in all brownfield locations.

3.4.2 Housing and Neighborhood Data

Housing transactions data are purchased from Dataquick Information Services through Duke University, and provide the universe of housing transactions in the state of Massachusetts from 1998 through 2011. The data contain detailed house attributes recorded from the most recent tax assessment, including square footage, age, the number of bedrooms and bathrooms, and longitude and latitude coordinates. Information about the history of transactions for each property includes the transaction price and date, the names of the buyers and sellers, as well as information about the buyer's mortgage loan.

The analysis limits transaction types to arms length transactions and properties that are owner-occupied. Observations that are missing in any of the attributes are dropped, as well as houses that have undergone major improvements after the beginning of the sample time frame. Houses that sold more than once per year or four times per the window of house sales (14 years) are excluded. Prices are normalized to January 2000 dollars using the monthly All Urban Consumer Price Index for Housing in the U.S. Northeast Region available from the Index (2012). The analysis then excludes the 1st and 99th percentile of the observed price distribution. As brownfields are fairly local disamenties, to ensure that they are capitalized into housing prices, houses are first mapped to the nearest brownfield site using a Geographical Information System (GIS), and then only houses within 3km of a brownfield site are retained. Summary statistics for the housing data are given in Panel A of Table 3.4.

District level data in Massachusetts are obtained from Mass.gov, the official state website. Districts are defined as areas where 'public education services [are provided] for the area's residents' (U.S. Census Bureau). We retrieve information on crime, school quality and a measure of poverty for the 89 school districts within 3km of brownfield sites.[5] Data describing exposure to crime are taken from the FBI Uniform Crime Reports. This gives the number of violent crimes as reported by police agencies (corresponding to towns/cities) in Massachusetts from 1998 to 2011. Crime levels are then divided by the population for each district over time to make the measure more comparable across geographic areas. As a measure of school quality, we use the percentage of grade 10 students in each school district that achieved a score of Advanced or Proficient in math on the Massachusetts Comprehensive Assessment System (MCAS) test. We take the percentage of students in each district that come from low income households to gauge the poverty level for each district. Both measures are available at the school district level from 1998 to 2011 from the Massachusetts Department of Education website. Summary statistics for the housing data are given in Panel B of Table 3.4.

Locations for aquifers in Massachusetts are retrieved from the Office of Geographic Information (MassGIS (2014)), and mapped to the nearest brownfield sites based on the longitude and latitude coordinates using GIS. Figure 3.2 outlines the aquifers in Massachusetts. Brownfields are considered exposed to aquifers if it is situated directly above any portion of the aquifer.

3.5 Results

Table 3.5 gives the estimates from the basic hedonic regression. Column (1) provides estimates that control for structural characteristics of the house as well as site characteristics. Column (2) adds controls for district and brownfield characteristics. Column (3) controls for unobserved, time-invariant differences across school districts by including district-level fixed effects. The parameter on *maxtox* is negative and significant until fixed effects for district are included, which alludes to the concern raised earlier that brownfield sites in more industrial areas may cause an overstate-

[5] There are a total of 224 school districts in the state of Massachusetts.

ment of the depressed effect of toxicity on housing prices. With a median *maxtox* concentration of 3.92 found across all assessments, the effect of toxicity on housing prices is economically insignificant.

The OLS and FE estimators also find that exceeding contaminant thresholds respectively lead to housing price *increases* of 7.39% and 6.36% after controlling for observable neighborhood attributes and district-level unobservables. This counter-intuitive result is likely due to unobserved factors that are correlated with threshold exceedance and price. Examination of house attributes by threshold exceedance (Table 3.6, Panel A), and district attributes by threshold exceedance (Table 3.6, Panel B) suggest that there are clearly significant differences between houses sold near sites where assessments revealed that contaminants exceed thresholds versus those that did not. In the lower sections of panels A and B, attributes are compared after demeaning by district to account for fixed effects.

Most differences in physical house characteristics are rejected, but to the extent that data on structural characteristics is fairly comprehensive, these can be controlled for. The significant differences in neighborhood attributes are more concerning as they are more likely sources of bias for the estimated parameter on *exceedthreshold*. The within-district means of % low income and % crime (neighborhood 'bads') are higher for areas where thresholds were not exceeded. If areas with fewer neighborhood 'bads' coincide with better neighborhood amenities, then, based on MassDEP regulations, one would find a higher instance of threshold exceedance, which may be the reason for the counter-intuitive positive estimate on the coefficient for *exceedthreshold*.

To deal with this endogeneity problem, we instrument for *exceedthreshold* using a dummy variable for whether a site is located over an aquifer, denoted $aquifer$. The relevance condition should be satisfied since the MA Contingency Plan specifies that different thresholds be used for contaminants if the site is located near an underground water source. Moreover, this can be tested in the first stage of the the IV regression. More difficult to justify is whether the location of the aquifer is uncorrelated with other determinants of housing prices, and that its only effect on housing prices is through increasing the chance of exceeding contaminant thresholds. We limit the sample of houses near brownfields that are within 500 meters of an aquifer.[6] Table 3.7 presents house (Panel A, top) and district (Panel B, top) attributes for brownfields by location on an aquifer. Significant differences by aquifer location exist in the number of bathrooms, percentage of condo and single family homes. These are likely due to the fact that brownfields, and the proximate houses, in more urban areas where there are more space constraints are less likely to locate on aquifers. To limit the comparison of houses that are more similar, we remove all condos in one of our instrumental variables specifications. Finally, note that by splitting the sample by aquifer rather than threshold exceedance the differences in neighborhood attributes are no longer significant.

First stage estimates using $aquifer$ as an instrument for thresholds are given in Column (1) of Table 3.9. The corresponding IV estimates are provided in Column (2). As expected, locating on an aquifer significantly increases the chances of exceeding contamination thresholds, satisfying the relevance condition for the instrument. The IV estimate finds that exceeding a threshold leads to a 10.8% *decline* in housing prices, and that after accounting for this threshold exceedance, increasing toxicity has a statistically insignificant effect. This coefficient is significant at the 5% level with standard errors clustered at the school district level. Finally, we remove all condos from our sample (approx. 20%), and present results in Column (3). The estimate on threshold exceedance becomes insignificant. This could be potentially concerning if places with condos in suburban Massachusetts are the residences that are serviced

[6]Table 3.10 give results that increase this distance. Generally, farther distances to aquifer causes the estimate on threshold exceedance to be insignificant. This is not surprising as aquifer location becomes less of a determinant in threshold exceedance for brownfields far from aquifers.

by piped water. In this case, since the magnitude of the estimate decreased, it would suggest that we have removed some of the correlation between aquifer location and PWSA, which would invalidate the instrument. Therefore, to have more definitive results, recovering data maps for PWSA is important for this paper going forward.

3.6 Conclusion

This paper examines if pollution information capitalizes into housing values differently depending on how it is conveyed. Since brownfield contamination can be disclosed as a continuous or discrete measure, and since the thresholds for the discrete measures vary in our context while holding contamination constant, we are able to test whether exceeding a threshold on contamination even matters after conditioning on the actual level of contamination. An additional complication arises with estimating the effect of threshold exceedance on housing prices: contaminant thresholds are endogenous as they depend on the proximity to institutions around the site that could affect housing values. We deal with this endogeneity problem by using aquifer location as an instrument to induce variation in the likelihood of exceeding toxin thresholds.

In our fixed effects specification that ignores threshold endogeneity, we find that exceeding thresholds leads to an increase in housing values of around 5% after controlling for house attributes, time-varying neighborhood attributes, brownfield attributes, and school district-level fixed effects. We also find that the actual amount of toxicity has a negative but insignificant effect on housing prices. This counter-intuitive result on threshold exceedance suggests that the fixed effect estimator that removes time-invariant unobserved factors at the district level does not adequately control for the local amenity differences at a finer geographic level that are also correlated with threshold exceedance. We examine the differences in attribute means conditional on our instrument (as opposed to threshold exceedance) to check if attributes are balanced across our instrument. After instrumenting for threshold exceedance with presence of an aquifer, our estimates indicate a 10.8% *decrease* in housing values from exceeding contaminant thresholds. The parameter on continuous toxicity values remains insignificant in the IV specification. Finally, the most obvious threat to the exogeneity of the instrument is whether aquifer presence is negatively correlated with a property's dependence on piped water systems, which would tend to depress house values holding other attributes constant. After controlling for piped water service areas in eastern Massachusetts, we do not find any changes in our estimates. Although it is encouraging that our estimate is robust to the inclusion of the piped water area dummy variable, until more complete data on piped water systems can be obtained, we cannot completely rule out this concern.

The findings in this paper suggest that information presented in a more interpretable manner (binary thresholds) are easier to process. More specifically, binary categories used to evaluate contamination is negatively capitalized into housing markets, whereas the effect of a continuous toxicity measure is statistically insignificant. Given this result, if one of the aims of the TRI program or the Right-to-Know act is to develop an informed public, then careful consideration of the format in which information is presented is important, and would play a large role in helping information provision policies to achieve that outcome.

3.7 Tables

Table 3.1: Chemical Group by Threshold Exceedance

	exceedthreshold		
Chemical Group	0	1	Total
Post Transition Metals	155	100	255
Transition Metals	707	179	886
Alkaline Earth & Alkali Metals	384	102	486
Other Elements & Asbestos	287	60	347
Aliphatics	444	154	598
Other Aromatics	947	222	1,169
Xylenes	225	29	254
3 ring PAH	661	65	726
4 ring PAH	326	119	445
5+ ring PAH	563	275	838
PCB & Oil	130	32	162
Hydrocarbons & Other organics	194	13	207
Organochlorides	580	185	765
Aromatic Organochlorides	185	21	206
Organobromides	95	4	99
Naphthalenes	295	39	334
Pesticides	101	14	115
Ketone	97	11	108
Aromatics with multiple substitent groups	125	19	144
Nitro containing compounds	88	8	96
Total	6,589	1,651	8,240

Table 3.2: List of Chemicals

Chemical	Freq.	Percent	Chemical	Freq.	Percent	Chemical	Freq.	Percent
1,1,1,2 Tetrachloroethane	11	0.13	Aroclor 1221	13	0.16	Fluoranthene	152	1.84
1,1,1 Tetrachloroethane	1	0.01	Aroclor 1232	13	0.16	Fluorene	147	1.78
1,1,1 Trichloroethane	52	0.63	Aroclor 1242	15	0.18	Freon 11	2	0.02
1,1,2,2 Tetrachloroethane	14	0.17	Aroclor 1248	19	0.23	Freon 113	3	0.04
1,1,2 Trichloroethane	15	0.18	Aroclor 1250	2	0.02	Freon 114	1	0.01
1,1,2 Trichlorotrifluoroethane	2	0.02	Aroclor 1254	28	0.34	Freon 12	2	0.02
1,1 Biphenyl	1	0.01	Aroclor 1260	32	0.39	Heptachlor	7	0.08
1,1 Dichloroethane	45	0.55	Aroclor 1262	9	0.11	Heptachlor Epoxide	3	0.04
1,1 Dichloroethene	24	0.29	Aroclor 1268	8	0.1	Heptane	1	0.01
1,1 Dichloroethylene	7	0.08	Arsenic	158	1.92	Hexachlorobenzene	7	0.08
1,1 Dichloropropene	7	0.08	Asbestos	3	0.04	Hexachlorobutadiene	15	0.18
1,2,3,4,6,7,8,9 OCDD	2	0.02	Azobenzene	3	0.04	Hexachlorocyclopentadiene	1	0.01
1,2,3,4,6,7,8,9 OCDF	2	0.02	Barium	138	1.67	Hexachloroethane	4	0.05
1,2,3,4,6,7,8 HpCDD	4	0.05	Benzene	130	1.58	Indeno (1,2,3 cd) pyrene	140	1.7
1,2,3,4,7,8,9 HpCDF	2	0.02	Benzo (a) anthracene	145	1.76	Iodomethane	1	0.01
1,2,3,4,7,8 HxCDD	2	0.02	Benzo (a) pyrene	145	1.76	Iron	18	0.22
1,2,3,4,7,8 HxCDF	2	0.02	Benzo (a,e) pyrene	1	0.01	Isophorone	8	0.1
1,2,3,6,7,8 HxCDD	2	0.02	Benzo (b) fluoranthene	146	1.77	Isopropanol	1	0.01
1,2,3,6,7,8 HxCDF	2	0.02	Benzo (e) pyrene	7	0.08	Isopropyl Ether	1	0.01
1,2,3,7,8,9 HxCDD	2	0.02	Benzo (g,h,i) perylene	130	1.58	Lead	178	2.16
1,2,3,7,8,9 HxCDF	1	0.01	Benzo (j,k) Fluoranthene	1	0.01	Magnesium	13	0.16
1,2,3,7,8 PeCDD	2	0.02	Benzo (k) fluoranthene	141	1.71	Manganese	19	0.23
1,2,3,7,8 PeCDF	2	0.02	Benzoic Acid	1	0.01	Mercury	125	1.52
1,2,3 Trichlorobenzene	16	0.19	Benzyl chloride	1	0.01	Methoxychlor	5	0.06
1,2,3 Trichloropropane	8	0.1	Beryllium	80	0.97	Methyl Ethyl Ketone	2	0.02
1,2,3 Trimethylbenzene	1	0.01	Bis (2 Chloroethoxy) Methane	4	0.05	Methyl Isobutyl Ketone	6	0.07
1,2,4 Trichlorobenzene	24	0.29	Bis (2 Chloroethy) Ether	5	0.06	Methyl tert butyl ether	131	1.59
1,2,4 Trimethylbenzene	75	0.91	Bis (2 Ethyhexyl) phthalate	15	0.18	Methylene Chloride	33	0.4
1,2 Dibromo 3 Chloropropane	8	0.1	Bis (2 Ethylhexyl) phthalate	2	0.02	Motor Oil	1	0.01
1,2 Dibromoethane	12	0.15	Bromobenzene	10	0.12	N Nitrosodiphenylamine	2	0.02
1,2 Dichlorobenzene	32	0.39	Bromochloromethane	10	0.12	Naphthalene	210	2.55
1,2 Dichloroethane	21	0.25	Bromodichloromethane	13	0.16	Nickel	102	1.24
1,2 Dichloropropane	15	0.18	Bromoform	14	0.17	Nitrate	1	0.01
1,2 Diphenylhydrazine	1	0.01	Bromomethane	16	0.19	Nitrite	1	0.01
1,2 dichloropropane	1	0.01	Butylbenzylphtalate	7	0.08	Nitrobenzene	4	0.05
1,3,5 Trimethylbenzene	69	0.84	C10 C28 Medium Petroleum Distillate	1	0.01	Pentachlorophenol	12	0.15
1,3 Butadiene	3	0.04	C11 C22 Aromatics	179	2.17	Perchloroethylene	1	0.01
1,3 Dichlorobenzene	26	0.32	C13 C16 Aliphatics	1	0.01	Perylene	5	0.06
1,3 Dichloropropane	7	0.08	C16 C36 Heavy Petroleum Distillate	1	0.01	Pesticides	3	0.04
1,4 Dichlorobenzene	38	0.46	C19 C22 Aromatics	1	0.01	Phenanthrene	161	1.95
1,4 Dioxane	8	0.1	C19 C36 Aliphatics	174	2.11	Phenol	7	0.08
1 Methylnaphthalene	12	0.15	C5 C10 Aromatics	2	0.02	Phosphorus	1	0.01
2,2 Dichloropropane	7	0.08	C5 C8 Aliphatics	127	1.54	Potassium	9	0.11
2,2 Oxybis (1 Chloropropane)	3	0.04	C6 C12 Light Petroleum Distillate	1	0.01	Propylene	1	0.01
2,3,4,6,7,8 HxCDF	2	0.02	C6 C36 Aromatics	2	0.02	Pyrene	154	1.87
2,3,4,7,8 PeCDF	2	0.02	C6 C8 Aliphatics	2	0.02	Pyridine	2	0.02
2,3,7,8 TCDD	2	0.02	C9 C10 Aromatics	128	1.55	Selenium	105	1.27
2,3,7,8 TCDD TEQ	1	0.01	C9 C12 Aliphatics	128	1.55	Silver	117	1.42
2,3,7,8 TCDF	2	0.02	C9 C18 Aliphatics	165	2	Sodium	9	0.11
2,4,5 Trichlorophenol	6	0.07	C9 C36 Aliphatics	1	0.01	Styrene	17	0.21
2,4,6 Trichlorophenol	5	0.06	Cadmium	135	1.64	Sulfate	1	0.01
2,4 Dichlorophenol	4	0.05	Calcium	9	0.11	Tetrachlorethene	64	0.78
2,4 Dimethylphenol	6	0.07	Carbazole	10	0.12	Tetrachloroethylene	18	0.22
2,4 Dinitrophenol	4	0.05	Carbon Disulfide	14	0.17	Tetrahydrofuran	8	0.1
2,4 Dinitrotoluene	4	0.05	Carbon Tetrachloride	13	0.16	Thallium	68	0.83
2,6 Dimethylnaphthalene	3	0.04	Chlordane	2	0.02	Toluene	147	1.78
2,6 Dinitrotoluene	4	0.05	Chlorobenzene	29	0.35	Trichloro fluoro methane	2	0.02
2 Butanone	26	0.32	Chlorodibromomethane	3	0.04	Trichloroethene	62	0.75
2 Chloronaphthalene	8	0.1	Chloroethane	20	0.24	Trichloroethylene	20	0.24
2 Chlorophenol	4	0.05	Chloroform	21	0.25	Trichlorofluoromethane	9	0.11
2 Chlorotoluene	9	0.11	Chloromethane	15	0.18	Vanadium	69	0.84
2 Hexanone	12	0.15	Chromium	151	1.83	Vinyl Acetate	5	0.06
2 Methylnaphthalene	159	1.93	Chrysene	146	1.77	Vinyl Chloride	42	0.51
2 Methylphenol	5	0.06	Cobalt	8	0.1	Xylene	92	1.12
2 Nitroaniline	2	0.02	Copper	50	0.61	Zinc	109	1.32
2 Nitrophenol	4	0.05	Cyanide	29	0.35	alpha BHC	4	0.05
3,3 Dichlorobenzidine	4	0.05	Cyclohexane	1	0.01	alpha Chlordane	4	0.05
3 Nitroaniline	2	0.02	Di n Butylphthalate	13	0.16	beta BHC	5	0.06
4,4' DDD	9	0.11	Di n Octylphthalate	6	0.07	bis (2 Chloroisopropyl) Ether	1	0.01
4,4' DDE	9	0.11	Dibenzo (a,h) anthracene	127	1.54	cis 1,2 Dichloroethene	45	0.55
4,4' DDT	11	0.13	Dibenzofuran	18	0.22	cis 1,2 Dichloroethylene	10	0.12
4,6 Dinitro 2 Methylphenol	2	0.02	Dibenzothiophene	2	0.02	cis 1,3 Dichloropropene	13	0.16
4 Isopropyltoluene	1	0.01	Dibromochloromethane	10	0.12	cis Dichloroethene	8	0.1
4 Bromophenyl phenylether	4	0.05	Dibromoethane	1	0.01	delta BHC	2	0.02
4 Chloro 3 Methylphenol	2	0.02	Dibromomethane	6	0.07	gamma BHC	3	0.04
4 Chloroaniline	4	0.05	Dichlorodifluoromethane	12	0.15	gamma Chlordane	6	0.07
4 Chlorophenyl Phenyl Ether	2	0.02	Dichloromethane	1	0.01	isopropylbenzene	54	0.66
4 Chlorotoluene	11	0.13	Dieldrin	11	0.13	m/p Cresol	1	0.01
4 Ethyltoluene	1	0.01	Diethyl Ether	4	0.05	m/p Xylene	86	1.04
4 Isopropylbenzene	1	0.01	Diethylphthalate	6	0.07	n Butylbenzene	49	0.59
4 Isopropyltoluene	23	0.28	Diisopropyl Ether	6	0.07	n Hexane	1	0.01
4 Methyl 2 Pentanone	9	0.11	Dimethylphthalate	4	0.05	n Nitroso di n Propylamine	2	0.02
4 Methylphenol	5	0.06	Dioxins	5	0.06	n Propylbenzene	59	0.72
4 Nitroaniline	2	0.02	Endosulfan	1	0.01	o Chlorotoluene	1	0.01
4 Nitrophenol	3	0.04	Endosulfan I	3	0.04	o Cresol	1	0.01
4 nitrophenol	1	0.01	Endosulfan II	3	0.04	o Xylene	76	0.92
Acenaphthene	140	1.7	Endosulfan Sulfate	1	0.01	p Chlorotoluene	1	0.01
Acenaphthylene	132	1.6	Endosulfan sulfate	1	0.01	p Isopropyltoluene	27	0.33
Acetone	40	0.49	Endrin	3	0.04	sec Butylbenzene	51	0.62
Acetophenone	5	0.06	Endrin Aldehyde	1	0.01	tert Amyl Methyl Ether	7	0.08
Aldrin	3	0.04	Endrin Ketone	4	0.05	tert Butylalcohol	1	0.01
Aluminum	9	0.11	Ethanol	1	0.01	tert Butylbenzene	19	0.23
Aniline	5	0.06	Ether	2	0.02	tert Butylethyl Ether	3	0.04
Anthracene	146	1.77	Ethyl Acetate	1	0.01	trans 1,2 Dichloroethene	21	0.25
Antimony	81	0.98	Ethyl Ether	1	0.01	trans 1,2 Dichloroethylene	9	0.11
Aroclor	5	0.06	Ethyl tert butyl Ether	3	0.04	trans 1,3 Dichloropropane	1	0.01
Aroclor 1016	13	0.16	Ethylbenzene	150	1.82	trans 1,3 Dichloropropene	12	0.15

Table 3.3: Brownfield & Assessment Summary Statistics

	Brownfield Attributes					
Variable	Obs.	Mean	Median	St. Dev.	Min.	Max.
Awarded	58	0.79	1.00	0.41	0.00	1.00
Proposal Score	56	94.13	96.00	14.14	53.00	118.00
Hazardous Substance	58	0.83	1.00	0.38	0.00	1.00
Petroleum	58	0.24	0.00	0.43	0.00	1.00
property size (acres)	54	4.49	1.55	6.29	0.02	27.00
	Assessment Results					
Variable	Obs.	Mean	Median	St. Dev.	Min.	Max
Samples Taken	262	31.45	28.00	25.79	1.00	162.00
maxconcentration	262	4461.94	3.92	32329.32	0.00	460000
Threshold Value	262	12947.91	300.00	80511.96	0.01	1000000
Soil	262	0.56	1.00	0.50	0.00	1.00
Groundwater	262	0.39	0.00	0.49	0.00	1.00
Other (air, surface water)	262.00	0.05	0.00	0.21	0.00	1.00

Table 3.4: Summary Statistics

			A. House Attributes			
Variable	Obs.	Mean	Median	St. Dev.	Min.	Max.
Price	290770	221647.30	189770.60	140491.30	21967.15	1166016.00
# Bathrooms	290770	1.77	1.50	0.89	0.00	90.00
# Bedrooms	290770	3.04	3.00	1.66	0.00	128.00
Sq. ft.	290770	1690.85	1447.00	1149.68	0.00	32410.00
Year Built	288215	1946.35	1953.00	42.32	1650.00	2012.00
Condo	290770	0.29	0.00	0.45	0.00	1.00
Single Family	290770	0.56	1.00	0.50	0.00	1.00
Age	288215	57.17	51.00	42.36	0.00	350.00
			B. Neighborhood Attributes (By School District)			
Variable	N	mean	p50	sd	min	max
% Low income	89	24.26	17.00	21.46	0.80	87.10
% Crime	88	0.00	0.00	0.00	0.00	0.02
% Proficient or Advanced	89	47.46	46.00	23.44	9.00	88.00
			C. Neighborhood Attributes (by Brownfield)			
Industrial activity nearby	58	0.45	0.00	0.50	0.00	1.00
Residential nearby	58	0.78	1.00	0.42	0.00	1.00
School nearby	58	0.59	1.00	0.50	0.00	1.00
Greenspace nearby	58	0.60	1.00	0.49	0.00	1.00

Table 3.5: Hedonic Price Regression

Dependent Variable: log(Price)	OLS		FE
exceedthreshold	-0.0165***	0.0739***	0.0636***
	(0.0032)	(0.0029)	(0.0031)
maxtox	-1.61e-10***	-4.04e-10***	-2.16E-11
	(0.0000)	(0.0000)	(0.0000)
Distance to Brownfield	8.22e-05***	5.87e-05***	7.45e-05***
	(0.0000)	(0.0000)	(0.0000)
# Bathrooms	0.184***	0.156***	0.130***
	(0.0018)	(0.0017)	(0.0015)
# Bedrooms	-0.0285***	-0.0162***	-0.00799***
	(0.0010)	(0.0009)	(0.0008)
Sq. Ft.	0.000135***	0.000119***	0.000119***
	(0.0000)	(0.0000)	(0.0000)
Year Built	0.0252***	-0.0509***	0.0122***
	(0.0011)	(0.0012)	(0.0013)
Age	0.0258***	-0.0508***	0.0112***
	(0.0011)	(0.0012)	(0.0013)
Condominium	0.175***	0.112***	0.143***
	(0.0043)	(0.0038)	(0.0034)
Single family	0.215***	0.0407***	-0.0928***
	(0.0052)	(0.0046)	(0.0042)
Awarded	0.345***	-0.0802***	-0.0375***
	(0.0052)	(0.0052)	(0.0067)
Proposal score	0.00476***	0.00725***	0.00552***
	(0.0001)	(0.0001)	(0.0002)
Hazardous substances	-0.303***	0.0950***	-0.118***
	(0.0071)	(0.0074)	(0.0200)
Petroleum	-0.417***	-0.126***	-0.118***
	(0.0067)	(0.0065)	(0.0174)
Property size (acres)	0.000101	-0.00293***	0.0175***
	(0.0003)	(0.0003)	(0.0007)
% Low income		0.00497***	0.000434
		(0.0001)	(0.0003)
% Crime		-7.912***	0.0995
		(0.4067)	(0.6570)
% Proficient or Advanced		0.0142***	-0.00254***
		(0.0001)	(0.0002)
Industrial activity nearby		-0.351***	-0.0928***
		(0.0030)	(0.0059)
Residential nearby		-0.0602***	-0.177***
		(0.0030)	(0.0045)
School nearby		0.0984***	0.0723***
		(0.0026)	(0.0040)
Greenspace nearby		0.0761***	0.0691***
		(0.0031)	(0.0043)
Constant	-39.91***	111.9***	-13.30***
	(2.1207)	(2.3697)	(2.5579)
Obs.	157,140	150,430	150,430
R-squared	0.332	0.506	0.358
House Characteristics	X	X	X
Year Fixed Effects	X	X	X
Neighborhood Characteristics		X	X
District Fixed Effects			X

Note: Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Attributes by Exceed Threshold

A. House Attributes	exceedthreshold = 1			exceedthreshold = 0					
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
Price	215403.15	123752.52	125710	195149.77	108658.53	29244	-25.77	0.00	Y
# Bathrooms	1.78	0.90	125710	1.77	0.91	29244	-1.61	0.11	N
# Bedrooms	3.10	1.69	125710	3.05	1.61	29244	-4.53	0.00	Y
Sq. Ft.	1703.18	1175.03	125710	1696.55	1067.49	29244	-0.88	0.38	N
Condominium	0.23	0.42	125710	0.21	0.41	29244	-5.99	0.00	Y
Single Family	0.61	0.49	125710	0.66	0.47	29244	13.77	0.00	Y
Age	56.30	41.03	124636	51.84	43.17	28815	-16.45	0.00	Y
	exceedthreshold = 1			exceedthreshold = 0					
Demean by District	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
Price	11812.42	100756.67	125710	7034.18	89251.27	29244	-7.46	0.00	Y
# Bathrooms	0.00	0.88	125710	0.01	0.89	29244	2.06	0.04	Y
# Bedrooms	-0.01	1.64	125710	-0.05	1.59	29244	-4.02	0.00	Y
Sq. Ft.	-3.28	1152.21	125710	-20.82	1050.88	29244	-2.38	0.02	Y
Condo	0.01	0.38	125710	0.02	0.39	29244	5.01	0.00	Y
Single Family	-0.01	0.44	125710	-0.02	0.45	29244	-4.96	0.00	Y
Age	1.12	38.14	124636	-0.10	40.82	28815	-4.80	0.00	Y
B. Neighborhood Attributes	exceedthreshold = 1			exceedthreshold = 0					
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
% Low income	28.74	23.17	614	25.12	22.44	136	-1.66	0.10	Y
% Crime	4.82E-03	4.32E-03	606	3.86E-03	3.94E-03	134	-2.38	0.02	Y
% Proficient or Advanced	59.97	19.48	614	60.85	21.48	136	0.47	0.64	N
	exceedthreshold = 1			exceedthreshold = 0					
Demean by District	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
% Low income	1.34	5.12	614	1.23	3.75	136	-0.25	0.80	N
% Crime	-2.38E-04	1.80E-03	606	1.13E-04	1.37E-03	134	2.13	0.03	Y
% Proficient or Advanced	7.76	13.78	614	7.30	13.12	136	-0.35	0.73	N

Table 3.7: Attributes by Whether Brownfield Located on Aquifer

A. House Attributes

	aquifer = 1			aquifer = 0					
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
Price	-356.27	86512.52	57387	308.65	84432.14	66241	1.37	0.17	N
# Bathrooms	-0.01	0.82	57387	0.01	0.99	66241	3.06	0.00	Y
# Bedrooms	0.00	1.62	57387	0.00	1.66	66241	0.85	0.39	N
Sq. Ft.	-2.74	1161.28	57387	2.37	1179.38	66241	0.76	0.44	N
Condo	0.01	0.36	57387	-0.01	0.38	66241	-5.85	0.00	Y
Single Family	0.00	0.43	57387	0.00	0.45	66241	3.21	0.00	Y
Age	0.05	40.43	56997	-0.04	38.59	65460	-0.38	0.70	N

B. Neighborhood Attributes

	aquifer = 1			aquifer = 0					
	Mean	St. Dev.	Obs.	Mean	St. Dev.	Obs.	t-stat	p-val	Reject?
% Low income	-0.03	4.11	258	0.08	4.48	285	-0.13	0.89	N
% Crime	3.10E-05	2.05E-03	254	-4.40E-05	2.12E-03	278	0.43	0.67	N
% Proficient or Advanced	-0.10	17.13	258	0.11	18.72	285	-0.23	0.82	N

Note: Variables are first demeaned by school district.

Table 3.8: Summary of Attribute Differences

	By Exceed Threshold				By Aquifer	
			Demean by School District		Demean by School District	
Characteristics	t-stat	Reject?	t-stat	Reject?	t-stat	Reject?
Price	**-25.77**	**Y**	**-7.46**	**Y**	1.37	N
# Bathrooms	-1.61	N	**2.06**	**Y**	**3.06**	**Y**
# Bedrooms	**-4.53**	**Y**	**-4.02**	**Y**	0.85	N
Sq. Ft.	-0.88	N	**-2.38**	**Y**	0.76	N
Condo	**-5.99**	**Y**	**5.01**	**Y**	**-5.85**	**Y**
Single Family	**13.77**	**Y**	**-4.96**	**Y**	**3.21**	**Y**
Age	**-16.45**	**Y**	**-4.80**	**Y**	-0.38	N
% Low income	**-1.66**	**Y**	-0.25	N	-0.13	N
% Crime	**-2.38**	**Y**	**2.13**	**Y**	0.43	N
% Proficient or Advanced	**0.47**	**Y**	-0.35	N	-0.23	N

Note: Rejection of the Null that differences in means across groups are equal are denoted in Bold.

Table 3.9: Hedonic Price Regression with Instrumental Variables

Dependent Variable:	First Stage exceedthreshold	IV log(Price)	IV Remove Condos log(Price)
exceedthreshold		-0.108**	-0.0770
		(0.0426)	(0.0564)
aquifer	0.849***		
	(0.0437)		
maxtox	-1.04e-10*	-3.79E-12	1.07E-11
	(0.0000)	(0.0000)	(0.0000)
Distance to Brownfield	2.40e-06	5.19e-05***	6.53e-05***
	(0.0000)	(0.0000)	(0.0000)
# Bathrooms	-0.00349	0.0896***	0.0709***
	(0.0023)	(0.0142)	(0.0126)
# Bedrooms	-5.11e-05	-0.00167	0.000372
	(0.0009)	(0.0078)	(0.0070)
Sq. Ft.	2.25e-07	0.000117***	0.000107***
	(0.0000)	(0.0000)	(0.0000)
Year Built	-0.0150	0.00617	0.00804
	(0.0181)	(0.0085)	(0.0080)
Age	-0.0151	0.00419	0.00583
	(0.0181)	(0.0086)	(0.0080)
Single Family	-0.00190	0.128***	0.0898***
	(0.0057)	(0.0362)	(0.0317)
Condominium	-0.00849	-0.228***	
	(0.0098)	(0.0661)	
Awarded	0.566***	0.0398	-0.0122
	(0.0683)	(0.0310)	(0.0306)
Proposal Score (std.)	-0.00204	0.00484***	0.00467**
	(0.0024)	(0.0015)	(0.0020)
Hazardous Substances	0.167**	0.00905	0.0318
	(0.0761)	(0.0440)	(0.0419)
Petroleum	0.0165	-0.0551**	-0.0103
	(0.0254)	(0.0257)	(0.0294)
Property size (acres)	0.0353***	0.00564	0.00293
	(0.0118)	(0.0052)	(0.0064)
% Low income	-0.00315	-0.00242	-0.00225
	(0.0062)	(0.0017)	(0.0014)
% Crime	-11.10	-6.119	-6.112
	(10.9047)	(4.5145)	(3.7841)
% Proficient or Advanced	0.00361	0.000401	-3.00e-05
	(0.0038)	(0.0012)	(0.0012)
Industrial activity nearby	0.541***	-0.0536	-0.0807
	(0.0878)	(0.0993)	(0.1135)
Residential nearby	-0.309***	0.0126	-0.0230
	(0.0482)	(0.0237)	(0.0284)
School nearby	-0.401***	0.0508*	0.0883***
	(0.0330)	(0.0251)	(0.0286)
Greenspace nearby	-0.391***	0.0173	-2.17e-05
	(0.0357)	(0.0263)	(0.0550)
Constant	27.21	-8.892	-4.594
	(27.6907)	(22.6718)	(15.7899)
Obs.	64,784	64,784	51,620
R-squared	0.416	0.606	0.612
House Characteristics	X	X	X
Year Fixed Effects	X	X	X
Neighborhood Characteristics	X	X	X
District Fixed Effects	X	X	X

Note: Cluster- Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Standard errors are clustered at the District level. There are 37 37 groups in the restricted sample.

Table 3.10: IV Regression using Different Aquifer Distances

	Distance to Aquifer $\leq$ B km			
VARIABLES	$B = 5$km	$B = 3$km	$B = 2$km	$B = 1$km
exceedthreshold	0.302	0.0810	-0.0471	-0.0448
	(0.4468)	(0.1610)	(0.0622)	(0.0849)
maxtox	0	0	-0	-0
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Distance to Brownfield	5.88e-05***	4.55e-05***	4.09e-05***	5.04e-05***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
# Bathrooms	0.111***	0.101***	0.0999***	0.0917***
	(0.0127)	(0.0112)	(0.0118)	(0.0123)
# Bedrooms	-0.00142	0.000865	0.000202	-0.000838
	(0.0070)	(0.0061)	(0.0062)	(0.0066)
Sq. Ft.	0.000111***	0.000107***	0.000102***	0.000108***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Year built	0.0164*	0.0125*	0.0103	0.00949
	(0.0088)	(0.0065)	(0.0072)	(0.0077)
Age	0.0148	0.0104	0.00811	0.00751
	(0.0090)	(0.0065)	(0.0072)	(0.0078)
Single family	0.0990***	0.0831**	0.0799**	0.0687*
	(0.0327)	(0.0317)	(0.0357)	(0.0389)
Condominium	-0.204***	-0.256***	-0.270***	-0.283***
	(0.0614)	(0.0511)	(0.0566)	(0.0639)
Awarded	0.0399	0.0901***	0.0882***	0.0691*
	(0.0872)	(0.0288)	(0.0260)	(0.0408)
Proposal Score (std.)	0.00351	0.00338	0.00127	0.00140
	(0.0054)	(0.0028)	(0.0013)	(0.0026)
Hazardous Substances	-0.107	-0.0805	-0.0543	-0.0905
	(0.1213)	(0.0829)	(0.0568)	(0.0588)
Petroleum	-0.0994**	-0.117***	-0.112***	-0.130***
	(0.0416)	(0.0342)	(0.0277)	(0.0373)
Property size (acres)	0.0323	0.0225	0.0106*	0.0106
	(0.0325)	(0.0140)	(0.0056)	(0.0102)
% Low income	-0.000714	-0.000881	-0.000854	-0.00110
	(0.0019)	(0.0024)	(0.0024)	(0.0026)
% Crime	-0.139	-1.904	-3.065	-2.584
	(5.8973)	(4.6599)	(4.8297)	(5.1672)
% Proficient or Advanced	-0.00435	-0.00237*	-0.00168*	-0.00129
	(0.0034)	(0.0013)	(0.0010)	(0.0011)
Industrial activity nearby	-0.165	-0.00982	-0.0417***	-0.0358
	(0.1316)	(0.0299)	(0.0149)	(0.1021)
Residential nearby	0.0512	0.121**	0.0895**	0.0754*
	(0.0628)	(0.0598)	(0.0421)	(0.0424)
School nearby	0.252	0.0829	0.0234	0.0350
	(0.2738)	(0.0857)	(0.0372)	(0.0429)
Greenspace nearby	-0.0746	0.0457	0.0265	0.0381
	(0.1053)	(0.0799)	(0.0427)	(0.0738)
Constant	-21.94	-13.96	-9.817	-7.466
	(18.1418)	(13.0196)	(14.4399)	(15.3910)
Obs.	128,584	116,431	106,149	85,174
R-squared	0.616	0.635	0.621	0.612

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

3.8 Figures

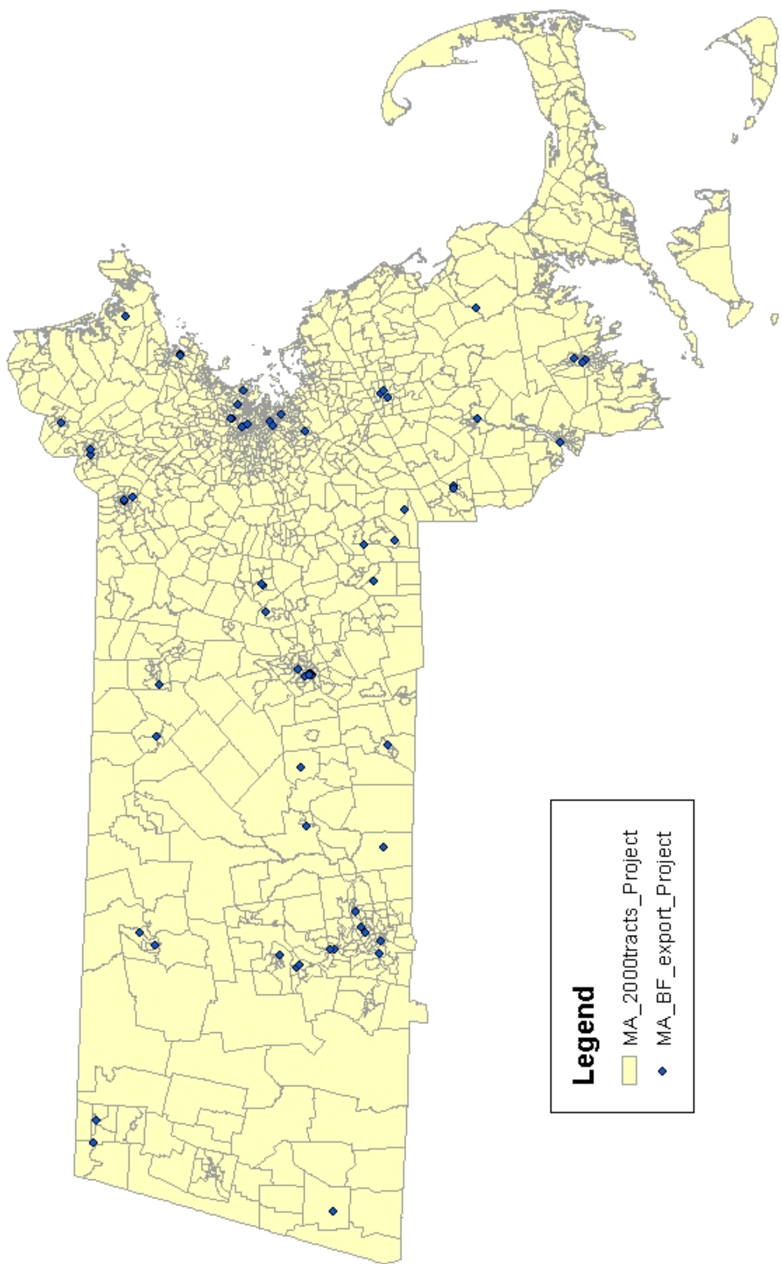


Figure 3.1: Brownfields in Massachusetts

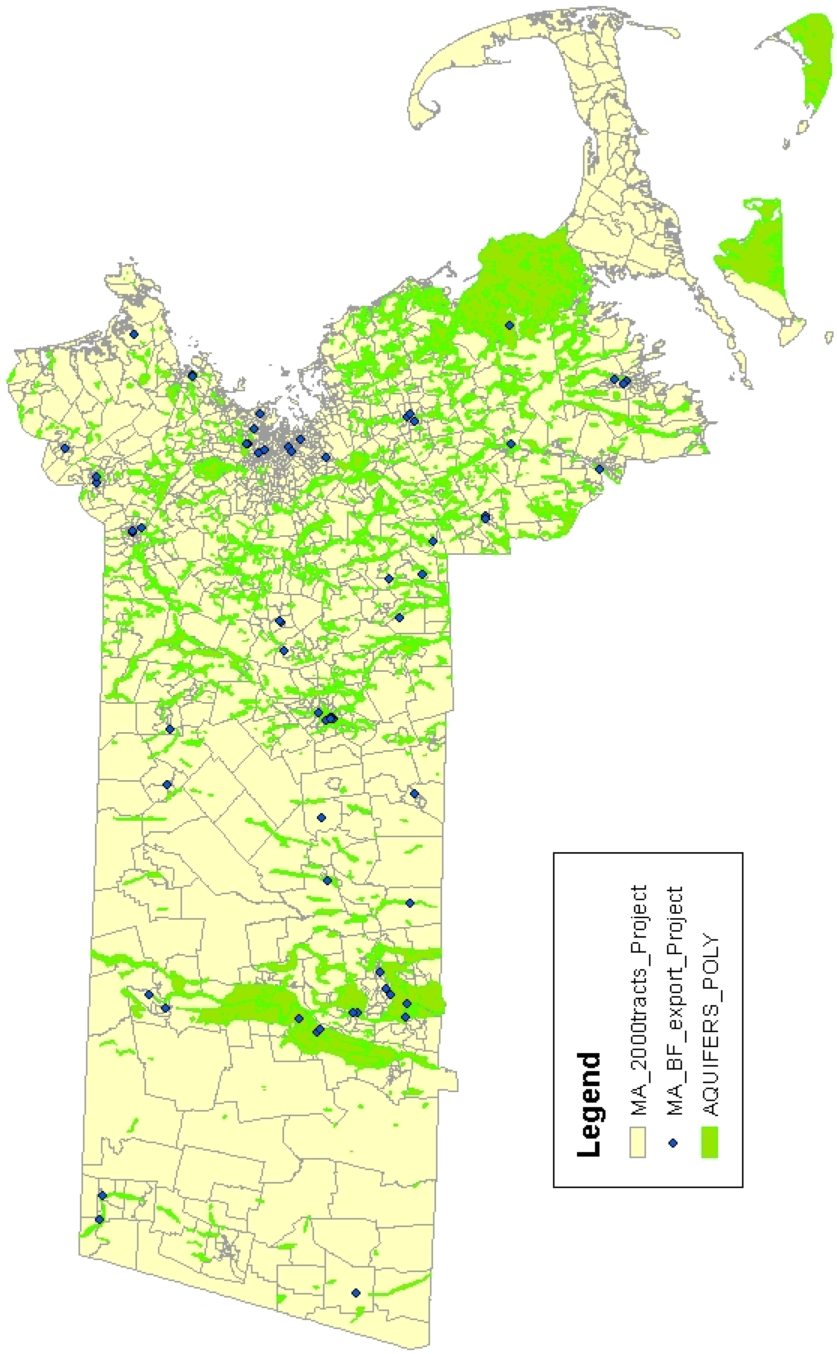


FIGURE 3.2: Aquifers in Massachusetts

Bibliography

Abadie, A. and Imbens, G. W. (2006), "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 74, 235–267.

Ackerberg, D. (2003), "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review*, 44, 1007–1040.

Adler, K. J., Anderson, R. C., Cook, Z., Dower, R. C., Ferguson, A. R., and Vickers, M. J. (1982), "The benefits of regulating hazardous waste disposal: land values as an estimator," *Washington, DC: Public Interest Economics Center.*

Arcidiacono, P. and Miller, R. (2011), "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79, 1823–1867.

Banerjee, A. and Duflo, E. (2009), "The Experimental Approach to Development Economics," *Annual Review of Economics*, 74, 151–178.

Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2011a), "A dynamic model of demand for houses and neighborhoods," Under Review.

Bayer, P., Geissler, C., and Roberts, J. (2011b), "Speculators and middlemen: The role of flippers in the housing market," *NBER Working Paper Series*, 16784.

Berry, S. T. (1994), "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, pp. 242–262.

Bieri, D. S., Kuminoff, N. V., and Pope, J. C. (2012), "The role of local amenities in the national economy," Working Paper.

Bishop, K. and Murphy, A. (2011), "Estimating the Willingness to Pay to Avoid Violent Crime: A Dynamic Approach," *The American Economic Review*, 101, 625–629.

Bishop, K. C. (2012), "A Dynamic Model of Location Choice and Hedonic Valuation," Working Paper.

Bishop, K. C. and Timmins, C. (2011), "Hedonic prices and implicit markets: Estimating marginal willingness to pay for differentiated products without instrumental variables," Tech. rep., National Bureau of Economic Research.

Brookshire, D., Thayer, M., Tschirhart, J., and Schulze, W. (1985), "A test of the expected utility model: Evidence from earthquake risks," *The Journal of Political Economy*, 93, 369–389.

Bui, L. T. and Mayer, C. J. (2003), "Regulation and capitalization of environmental amenities: Evidence from the toxic release inventory in Massachusetts," *Review of Economics and statistics*, 85, 693–708.

Chan, T. and Hamilton, B. (2006), "Learning, private information, and the economic evaluation of randomized experiments," *Journal of Political Economy*, 114, 997–1040.

Chay, K. and Greenstone, M. (2005), "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113.

Chernew, M., Gowrisankaran, G., and Scanlon, D. (2008), "Learning and the value of information: Evidence from health plan report cards," *Journal of Econometrics*, 144, 156–174.

Decker, C. S., Nielsen, D. A., and Sindt, R. P. (2005), "Residential Property Values and Community Right-to-Know Laws: Has the Toxics Release Inventory Had an Impact?" *Growth and Change*, 36, 113–133.

DeGroot, M. H. (1970), *Optimal statistical decisions*, McGraw-Hill Company, New York, St Louis, San Francisco, London, Sydney, Toronto, Mexico, Panama.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38.

EPA, U. (2012), "Brownfield and Land Revitalization," Available at `http://www.epa.gov/brownfields`, (assessed April 2013).

EPA, U. (2014), "Emergency Planning and Community Right-to-Know Act (EPCRA)," Available at `http://www2.epa.gov/epcra-tier-i-and-tier-ii-reporting/what-epcr`, (assessed March 2014).

Fan, J. and Gijbels, I. (1996), *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, vol. 66, CRC Press.

Flavin, M. and Yamashita, T. (2002), "Owner-occupied housing and the composition of the household portfolio," *The American Economic Review*, 92, 345–362.

Gamper-Rabindran, S., Mastromonaco, R., and Timmins, C. (2011), "Valuing the benefits of superfund site remediation: Three approaches to measuring localized externalities," Tech. rep., National Bureau of Economic Research.

Gawande, K. and Jenkins-Smith, H. (2001), "Nuclear waste transport and residential property values: estimating the effects of perceived risks," *Journal of Environmental Economics and Management*, 42, 207–233.

Gayer, T. and Viscusi, K. W. (2002), "Housing price responses to newspaper publicity of hazardous waste sites," *Resource and Energy Economics*, 24, 33–51.

Gayer, T., Hamilton, J., and Viscusi, W. (2000), "Private values of risk tradeoffs at superfund sites: housing market evidence on learning about risk," *Review of Economics and Statistics*, 82, 439–451.

Graff Zivin, J. and Neidell, M. (2009), "Days of haze: Environmental information disclosure and intertemporal avoidance behavior," *Journal of Environmental Economics and Management*, 58, 119–128.

Graff Zivin, J., Neidell, M., and Schlenker, W. (2011), "Water quality violations and avoidance behavior: Evidence from bottled water consumption," in *American Economic Review: Papers and Proceedings*, vol. 101, pp. 448–453.

Hallstrom, D. and Smith, V. (2005), "Market responses to hurricanes," *Journal of Environmental Economics and Management*, 50, 541–561.

Hamilton, J. T. (1995), "Pollution as news: media and stock market reactions to the toxics release inventory data," *Journal of environmental economics and management*, 28, 98–113.

Haninger, K., Ma, L., and Timmins, C. (2012), "Estimating the Impacts of Brownfield Remediation on Housing Property Values," Nicholas Institute Working Paper No. EE 12-08.

Himmelberg, C., Mayer, C., and Sinai, T. (2005), "Assessing high house prices: Bubbles, fundamentals, and misperceptions," Tech. rep., National Bureau of Economic Research.

Hotz, V. and Miller, R. (1993), "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 60, 497–529.

Index, B. C. P. (2012), "All Urban Consumers (Current Series)," Available at `http://www.bls.gov/cpi/data.htm`, (assessed April 2013).

James, J. (2011), "Ability matching and occupational choice," *Federal Reserve Bank of Cleveland*.

Khanna, M. and Damon, L. A. (1999), "EPA's voluntary 33/50 program: impact on toxic releases and economic performance of firms," *Journal of environmental economics and management*, 37, 1–25.

Kiel, K. A. (1995), "Measuring the impact of the discovery and cleaning of identified hazardous waste sites on house values," *Land Economics*, pp. 428–435.

Kohlhase, J. E. (1991), "The impact of toxic waste sites on housing values," *Journal of urban Economics*, 30, 1–26.

Konar, S. and Cohen, M. A. (1997), "Information as regulation: the effect of community right to know laws on toxic emissions," *Journal of environmental Economics and Management*, 32, 109–124.

Kuminoff, N. and Pope, J. (2011), "'Rosen Bias' in the Willingness to Pay for Nonmarket Goods," Under Review.

Kuminoff, N. V., Smith, V. K., and Timmins, C. (2013), "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets," *Journal of Economic Literature*, 51, 1007–1062.

Linden, L. and Rockoff, J. E. (2008), "Estimates of the impact of crime risk on property values from Megan's laws," *The American Economic Review*, pp. 1103–1127.

Ma, L. (2014), "Learning in a Hedonic Framework: Valuing Brownfield Remediationa," Working Paper.

MassGIS (2014)Available at `http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis`, (assessed March 2014).

Mastromonaco, R. (2011), "A Dynamic General Equilibrium Analysis of the Equity Implications and Welfare Effects of Dis-proportionate Exposure," Working Paper.

McClelland, G. H., Schulze, W. D., and Hurd, B. (1990), "The Effect of Risk Beliefs on Property Values: A Case Study of a Hazardous Waste Site1," *Risk analysis*, 10, 485–497.

McCluskey, J. J. and Rausser, G. C. (2003), "Hazardous waste sites and housing appreciation rates," *Journal of Environmental Economics and Management*, 45, 166–176.

Michaels, R. G. and Smith, V. K. (1990), "Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites," *Journal of Urban Economics*, 28, 223–242.

Oberholzer-Gee, F. and Mitsunari, M. (2006), "Information regulation: Do the victims of externalities pay attention?" *Journal of Regulatory Economics*, 30, 141–158.

of Environmental Protection, M. D. (2009), "Massachusetts Contingency Plan," Available at `http://www.mass.gov/dep/cleanup/laws/regulati.htm`, (assessed April 2013).

Palmquist, R. B. (2005), "Property value models," *Handbook of environmental economics*, 2, 763–819.

Pope, J. (2008), "Fear of crime and housing prices: Household reactions to sex offender registries," *Journal of Urban Economics*, 64, 601–614.

Poterba, J. M., Weil, D. N., and Shiller, R. (1991), "House price dynamics: The role of tax policy and demography," *Brookings Papers on Economic Activity*, 1991, 143–203.

Powers, N. E. (2013), "Measuring The Impact Of The Toxics Release Inventory: Evidence From Manufacturing Plant Births," Tech. rep.

Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition," *The Journal of Political Economy*, 82, 34–55.

Rust, J. (1987), "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

Sanders, N. J. (2013), "Toxic Assets: How the Housing Market Responds to Environmental Information Shocks," Under Review.

Schulze, W., McClelland, G., Hurd, B., and Smith, J. (1986), "Improving Accuracy and Reducing Costs of Environmental Benefit Assessments," *vol. IV: A Case Study of a Hazardous Waste Site: Perspectives from Economies and Psychology. Draft Report, US Environmental Protection Agency, Contract CR812054-01-1.*

Siniavskaia, N. (2011), "Property Tax Rates by County and City," .

Smith, V. K. and Johnson, F. R. (1988), "How do risk perceptions respond to information? The case of radon," *The Review of Economics and Statistics*, 70, 1–8.

Taylor, L. O. (2003), "The hedonic method," in *A primer on nonmarket valuation*, pp. 331–393, Springer.

Viscusi, W. K., Magat, W. A., and Huber, J. (1986), "Informational regulation of consumer health risks: an empirical evaluation of hazard warnings," *The Rand Journal of Economics*, pp. 351–365.

Biography

Lala Xun Ma was born on August 31, 1985 in Beijing, China. At the age of three, her family relocated to Massachusetts where she spent most of her childhood and graduated from Belmont High School in 2003. She attended Tufts University in Medford, MA, and graduated *Magna Cum Laude* in 2007 with Bachelor of Arts degrees in Economics and Mathematics. She also received a Masters of Arts in Economics from Tufts in the same year. While at Tufts, she was the president of the Chinese Student Association and was an active volunteer in the Tufts Big Sister-Little Sister program. After graduation, she worked as a consultant at IBM in Washington, DC for two years. In 2009, she began the PhD program in economics at Duke University in Durham, North Carolina. Lala will join the Gatton College of Business and Economics at the University of Kentucky as an assistant professor in the fall of 2014.